

# TECHNOLOGY, MIND & SOCIETY

AN APA CONFERENCE

**OCTOBER 3-5**  
**WASHINGTON, DC**  
GRAND HYATT WASHINGTON



AMERICAN  
PSYCHOLOGICAL  
ASSOCIATION

2019 PROGRAM

**#APATech19**

# TECHNOLOGY, MIND & SOCIETY



AMERICAN  
PSYCHOLOGICAL  
ASSOCIATION

## Full Program

### TABLE OF CONTENTS

Senior Program Committee.....	4
Review Committee.....	4
Sponsor Acknowledgments.....	5
Conference Exhibitors.....	6
Agenda at a Glance.....	7
<b>Program .....</b>	<b>10</b>
THURSDAY .....	10
FRIDAY .....	13
SATURDAY.....	96
Grand Hyatt Washington.....	162

#### THANK YOU TO OUR SPONSORS



RECOVERYPOINT

#### IN COOPERATION WITH

Association for the Advancement of Artificial Intelligence  
The Center for Innovative Research in Cyberlearning  
APA Division 3: Society for Experimental Psychology and Cognitive Science  
APA Division 13: Society of Consulting Psychology  
APA Division 14: Society for Industrial and Organizational Psychology  
APA Division 21: Applied Experimental and Engineering Psychology  
APA Division 46: Society for Media Psychology and Technology

## WIFI

**SSID:** HYATTMEETING

**Password:** TECH2019

*Case Sensitive*

## ADA/ACCESSIBILITY INFORMATION

Please see APA staff for assistance.

## GET THE MOST OUT OF TMS 2019

Set up your attendee account, view the full agenda, and start building a personal schedule of the presentations you want to attend.

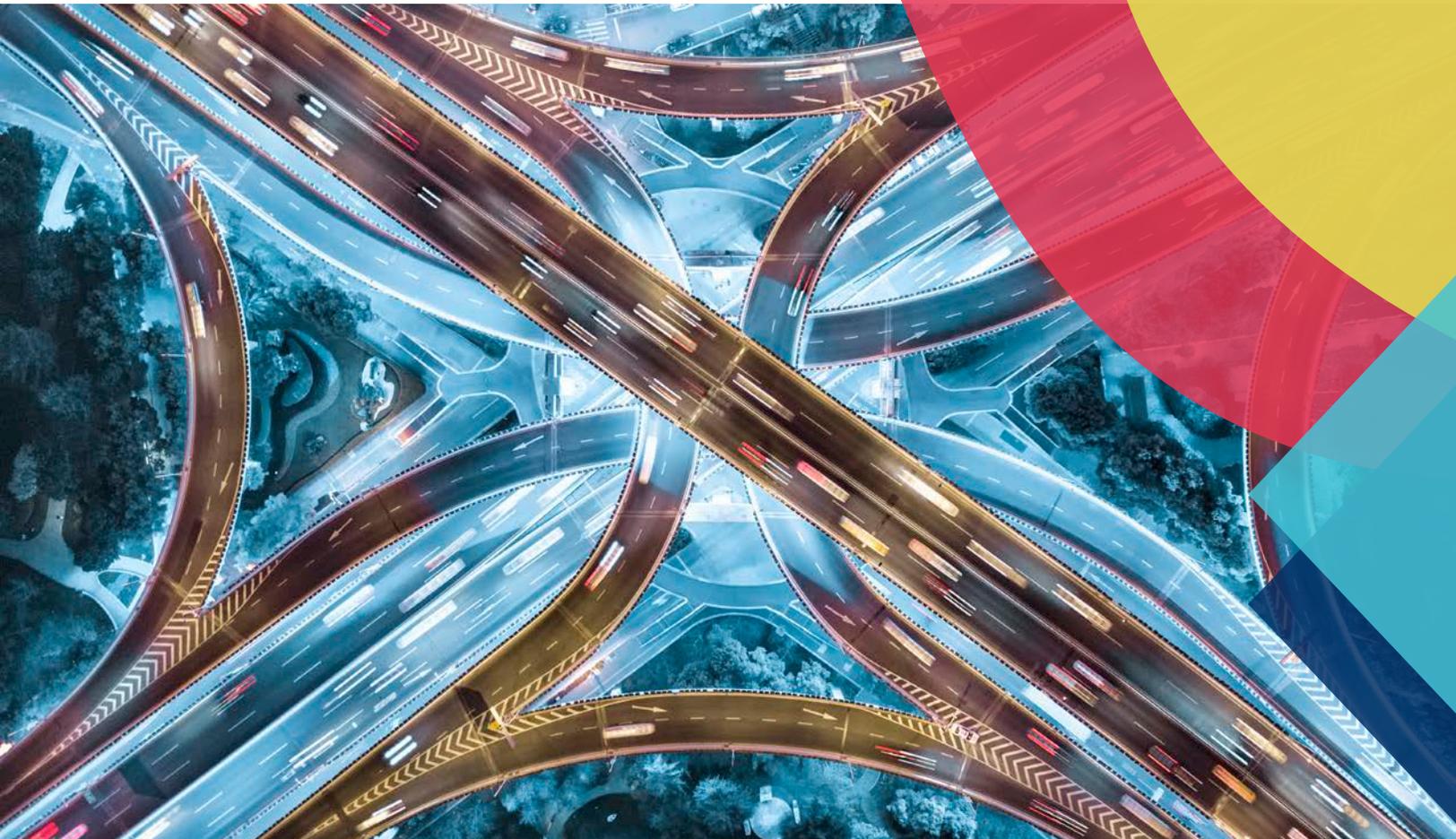
## TMS2019.HUBB.ME

## FOLLOW APA ON SOCIAL MEDIA

FACEBOOK: @AmericanPsychologicalAssociation

TWITTER: @APA

LINKEDIN: /company/american-psychological-association



## SENIOR PROGRAM COMMITTEE

**David W. Ballard**, *Conference Chair*  
American Psychological Association

**Tara Behrend**, *Committee Co-Chair*  
George Washington University

**H. Chad Lane**, *Committee Co-Chair*  
University of Illinois, Urbana-Champaign

**Tim Bickmore**  
Northeastern University

**Fran Blumberg**  
Fordham University

**Walter (Wally) Boot**  
Florida State University

**Will Browne**  
Victoria University of Wellington

**Angelo Cangelosi**  
University of Manchester

**Jing Chen**  
Old Dominion University

**Deborah Forster**  
University of California San Diego

**Jeff Hancock**  
Stanford University

**Peter Hancock**  
University of Central Florida

**Nicole Krämer**  
University of Duisburgh-Essen

**Jerry Krueger**  
Krueger Ergonomics Consultants

**Dar Meshi**  
Michigan State University

**Dan Morrow**  
University of Illinois Urbana-Champaign

**Anat Rafaeli**  
Technion Israel Institute of Technology

**Florian Röhrbein**  
Technical University of Munich

**Juliana Schroeder**  
University of California Berkeley

**Kim-Phuong L. Vu**  
California State University Long Beach

**Adrian Ward**  
University of Texas Austin

## REVIEW COMMITTEE

Sebiha Balci  
David Ballard  
Jaime Banks  
Heather Barker  
Barbara Barry  
David Bilkey  
Ramakrishna Biswal  
Karen Black  
Emmanuel Blanchard  
Fran Blumberg  
Meyran Boniel-Nissim  
Justin Bonny  
Walter Boot  
Nigel Bosch  
Nick Bowman  
Andrea Bublitz  
Stephanie Buono  
Kaitlyn Burnell  
Sandra Calvert  
dymna casey  
Marc Cavazza  
Daniela Conti  
Sara Czaja  
Paul Derby  
Jon Elhai  
Jonas Frich  
Sharath Chandra Guntuku  
Rachelle Hippler  
Bruce Homer  
Elizabeth Skora Horgan  
Matt Howard  
Erica Kleinknecht  
Jeff Krichmar  
Helianthe Kort  
Edward Kravitz  
Yi-Ching Lee  
Laura Leventhal  
Sunny Liu  
Flora Ma  
Leanne Ma  
Athanasios Mazarakis  
Hannah Mieczkowski  
Ali Momen  
Teresa Ober  
Fred Oswald  
Eva Parkhurst  
Susan Persky  
Eli Pincus  
Roshni Raveendhran  
Lisa Renzi-Hammond

Felice Resnik  
Ashley Ricker  
Ganit Richter  
Ma. Mercedes Rodrigo  
Katharina Rohlfing  
Ronny Scherer  
Juliana Schroeder  
Keke Schuler  
Matthew Smith  
Thitirat Sriplo  
Taylor Stephens  
Jeremiah Still  
Ilaria Torre  
Kim Vu  
Staci Weiss  
Claudia Wenzel  
Lorin Williams  
Ying Xu  
Zhenhua Xu  
Caglar Yildirim  
Michael Zimmer  
Jennifer Zosh

## SPONSOR ACKNOWLEDGMENTS

We would like to thank our sponsors for helping to make the Technology, Mind, and Society conference possible.



---

### **PLAN B TECHNOLOGIES, INC.**

185 Admiral Cochrane Dr., Suite 150  
Annapolis, MN 21401  
[www.planbtech.net](http://www.planbtech.net)

Plan B Technologies, Inc. (PBT) is an IT solutions provider with expertise in storage, backup, recovery and archiving, security, networking, Microsoft, cloud and virtualization solutions. PBT delivers the complete lifecycle of professional services including project management, consultation, assessment, design, deployment, support, and staff augmentation. Professional services are delivered by engineers who are cross-trained with over 300 technical certifications. PBT listens carefully to every customer, striving to deliver the right solution—unbiased by product—enabling our customers to achieve their business objectives, stability and success.

## RECOVERYPOINT

---

### **RECOVERYPOINT**

75 West Watkins Mill Rd.  
Gaithersburg, MD 20878  
[www.recoverypoint.com](http://www.recoverypoint.com)

Recovery Point is a leading national provider of IT infrastructure and business resilience services. We support all technology environments, from the mainframe to the desktop. Our comprehensive solutions include compliant private, multitenant, hyperscale, and hybrid cloud offerings, certified hosting infrastructure, national network services and a broad spectrum of cloud-based business resilience services including infrastructure as a service (IaaS), disaster recovery as a service (DRaaS), and backup as a service (BaaS). Recovery Point's unique integrated service delivery model reduces risk, eliminates vendor fragmentation and accelerates recovery processes by providing the complete range of resources required to enable survival of catastrophic events within all-inclusive, highly secure facilities.

## EXHIBITORS

### BRAINTRAIN, INC.

727 Twin Ridge Ln.  
North Chesterfield, VA 23235  
[www.braintrain.com](http://www.braintrain.com)

BOOTH 15

BrainTrain invites you to discover the NEW SCoPES Test Battery for FREE. Using SCoPES researchers can evaluate 20+ cognitive skills and create customized test batteries selecting from its 150+ different tests. PLUS you can now easily and accurately measure an individual's brain wave activity while they take the test. All EEG data is automatically de-artifacted and saved for each test. A fully functioning FREE trial of SCoPES is available upon request. We are also looking to work with researchers in joint projects that use SCoPES. Email [info@braintrain.com](mailto:info@braintrain.com) or call 800-822-0538 x105 for more information and a FREE trial.

### LIMBIX

839 Emerson Street  
Palo Alto, CA 94301  
[limbix.com](http://limbix.com)

BOOTH 13

Limbix creates prescription digital therapeutics for mental health disorders using mobile apps and virtual reality. The company currently provides the Limbix VR Kit to mental health professionals around the world so providers can offer patients mindfulness training, exposure therapy, psycho-education, and more.

### MOVIA ROBOTICS

72 Prospect Place  
Bristol, CT 06010  
[www.moviarobotics.com](http://www.moviarobotics.com)

BOOTH 14

Movia Robotics builds collaborative robotic systems for companionship, education and therapeutic uses. Our autonomous and semi-autonomous systems provide robot assisted instruction for special needs and general education. Our systems provide automatic data collection and analytics. We are building tools that extend the capability of teachers and therapists, helping them help people who have special needs. Our hope is to make a positive impact and change the lives of children, adults and their families. Please contact [info@moviarobotics.com](mailto:info@moviarobotics.com) for more information..

### ROSE: SMARTER MENTAL HEALTH

80 M Street, SE  
Washington, DC 20003  
[www.askrose.com](http://www.askrose.com)

BOOTH 4

Rose is a HIPAA-compliant, technology-driven mental health platform designed to enhance today's therapy process. We have built a user-friendly, real-time web-, mobile-, and desktop-based application that allows patients and therapists to interface with multiple touchpoints, enabling clinicians to gain a more informed view of their patients' treatment progress. We leverage an online journaling feature that detects patients' semantics and can pick up affective disorders, and integrate daily mood and anxiety assessments to gauge patients' mental well-being. Our machine learning platform helps therapists detect symptoms, tailor individualized treatment plans, provide curated insights for self-care, and utilize multiple touchpoints with their patients to increase the efficiency of therapy sessions. Rose uses an evidence-based approach to mental health, but it does not aim to replace therapy or any single aspect of the mental health care process, but rather acts as a comprehensive solution to mental health care that encompasses a variety of treatment processes. Our mission is to transform the way mental health care is delivered by increasing the engagement between therapists and their patients. For more information visit us at [askrose.com](http://askrose.com)!

### TECHNOLOGY, MIND, AND BEHAVIOR JOURNAL

750 First St., NE  
Washington, DC 20002-4242  
[apa.org/pubs/journals](http://apa.org/pubs/journals)

BOOTH 6

*Technology, Mind, and Behavior* is a new, open access journal published by the American Psychological Association, is open for submissions! Helmed by founding editor Danielle S. McNamara, PhD., the journal will publish original experimental research and review articles in the area of human-technology interaction. As technology evolves, so do we. The era of "move fast and break things" is over. Many protest against issues like privacy violations, surveillance, economic inequities, and cyberbullying, all amplified by technology. We hope to publish research to illuminate the ways in which individuals adapt to these issues and consider the impact technology has on society and behavior. *Technology, Mind, and Behavior* is a "pause" button for engineers, technologists, and psychologists to examine what happens when we press "play." The journal will feature Gold Open Access articles, published after peer review from experts across the field.

### THERAPYSITES

3010 Highland Pkwy., Suite 200  
Downers Grove, IL 60515  
[www.therapysites.com](http://www.therapysites.com)

BOOTH 16

Most prospective therapy patients begin their search online. Can they find you? At TherapySites, we're passionate about the mental health community and connecting you to your patients. We bundle all the necessary features and services into one all-inclusive, cost-effective and stress-free package. TherapySites offers it all—responsive designs, online appointment requests, SEO, on-site credit card acceptance, print marketing, and more. Plus, all new websites come with a listing in *Psychology Today* for 6 months, completely free! Don't just get a website. Get the full package. TherapySites—websites for therapists, made simple.

### THOUGHT TECHNOLOGY, LTD.

5250 Ferrier, Suite 812  
Montreal, QC H4P 1L3, Canada  
[www.thoughttechnology.com](http://www.thoughttechnology.com)

BOOTH 5

Combining in-clinic sessions and home training, the first product that links TPS wireless sensors and Biograph Infiniti providing elaborate statistics/modifiable protocols not capable with a simple app. Light-weight in design, eVu TPS a portable sensor brings biosignal data measurement in the clinic and into the client's smartphone/tablet allowing physician monitored compliance/progress. Help clients help themselves with TTL's clinically relevant medical software/apps: Revolutionize clinical training.

### UNIVERSITY OF UTAH, PHYSICAL MEDICINE AND REHABILITATION

30 N. 1900 E., Room 1B620  
Salt Lake City, UT 84132  
[medicine.utah.edu/pmr](http://medicine.utah.edu/pmr)

BOOTH 7

The Autonomy & Performance Technology (APT) Lab with the Division of Physical Medicine and Rehabilitation at the University of Utah is dedicated to accelerating equity and inclusion in function and self-determination for the disabled community. Partnered with the Therapeutic Games & App (GApp) Lab and TetrAdapt we examine, build, and transfer a range of rehabilitation technologies. We will feature a breadth of our projects addressing acute and long-term rehabilitation for a range of needs, including cognition, comorbidity reduction, pain, and adaptive recreation.

# Agenda at a Glance

## THURSDAY

5:00-7:15 p.m. <b>Keynote</b> <i>Independence Ballroom</i>	Opening Session, Keynote, and Poster Highlights <b><u>Is Clinical Virtual Reality Ready for Primetime?</u></b> <i>Albert "Skip" Rizzo, PhD</i> 
7:15-8:30 p.m. <i>Constitution Ballroom</i>	Welcome Reception

## FRIDAY

7:30-9:00 a.m. <i>Constitution Ballroom</i>	Poster Session 1 and Continental Breakfast			
9:15-10:30 a.m. <b>Keynote</b> <i>Independence Ballroom</i>	<b><u>Aging and Technology Systems: Benefits and Challenges</u></b> <i>Sara Czaja, PhD</i> 			
10:45 a.m.- Noon <b>Concurrent Sessions 1</b>	Games and Virtual Reality For Fostering Education and Knowledge <i>Paper Session. Lafayette Park.</i>	Bias and Technology <i>Paper Session. Farragut Square.</i>	Using Digital Data to Measure and Understand Mental Health <i>Symposium. Independence FG.</i> 	Using Online Technologies to Build Computing Skills That Meet Future Workforce Needs <i>Panel. Independence D.</i>
	Wearable Biometrics Technology: Investigating its use in simulated and live law enforcement training <i>Symposium. Independence BC.</i> 	Developing Technology-Based Mental Health Interventions <i>Paper Session. Independence HI.</i> 	Robotics and Psychology <i>Paper Session. Independence E.</i>	
Noon-1:15 p.m.	Lunch on your own			
1:15-2:30 p.m. <b>Concurrent Sessions 2</b>	Interactivity with Automation <i>Paper Session. Independence D.</i>	Older Adults and Technology <i>Paper Session. Independence BC.</i>	Games for Growth: Innovative Approaches to Applying Digital Games in Therapy <i>Panel. Independence HI.</i> 	Machine Learning For Analyzing Behavior <i>Paper Session. Lafayette Park.</i> 

**Earn CE credits at TMS 2019:** Select sessions at this conference have been reviewed and approved by the American Psychological Association's (APA) Office Continuing Education in Psychology (CEP) to offer Continuing Education (CE) credit for psychologists. Full attendance is required at each session for which you are claiming CE credit. Partial credit is not awarded. The CEP Office maintains responsibility for the delivery of the session.

## AGENDA AT A GLANCE

**Earn CE credits at TMS 2019:** Select sessions at this conference have been reviewed and approved by the American Psychological Association's (APA) Office Continuing Education in Psychology (CEP) to offer Continuing Education (CE) credit for psychologists. Full attendance is required at each session for which you are claiming CE credit. Partial credit is not awarded. The CEP Office maintains responsibility for the delivery of the session.

### FRIDAY CONTINUED

1:15–2:30 p.m. <b>Concurrent Sessions 2</b>	Rethinking Anthropomorphism: The Antecedents, Unexpected Consequences, and Potential Remedy for Perceiving Machines as Humanlike <i>Symposium.</i> <i>Independence E.</i>	Virtual Reality in The Tech Innovation Network—A Model of Clinical, Industry, and Academic Partnerships for Developing, Testing, and Implementing Mental Health Technologies <i>Panel.</i> <i>Independence FG.</i> 	National Science Foundation Funding Initiatives and Opportunities: The Role of Social, Behavioral, and Economic Sciences in Research on Human-Technology Interaction <i>Farragut Square.</i>	
2:45–4:00 p.m. <b>Concurrent Sessions 3</b> <i>All Paper Sessions</i>	Psychology and Cybersecurity <i>Farragut Square.</i>	The Future of Work, Job Skills, and Automation <i>Independence BC.</i> 	Psychological Assessment and Treatment Via Technology— <i>Independence HI.</i> 	Physiology and Psychology <i>Lafayette Park.</i>
	Algorithms and Advice <i>Independence E.</i>	Virtual Reality for Health and Well-Being <i>Independence FG.</i> 	Deception and Trust Online <i>Independence D.</i>	
4:00–4:30 p.m. <i>Independence Foyer</i>	Networking Break			
4:30–5:45 p.m. <b>Keynote</b> <i>Independence Ballroom</i>	<b><u>The Potential of Policy, Partnerships, and Combinatorial Innovation</u></b> <i>Kumar Garg, JD</i>			

### SATURDAY

7:30–9:00 a.m. <i>Constitution Ballroom</i>	Poster Session 2 and Continental Breakfast			
9:15–10:30 a.m. <b>Keynote</b> <i>Independence Ballroom</i>	<b><u>Stealth Assessment — What, Why, and How?</u></b> <i>Valerie Shute, PhD</i> 			
10:30–11:00 a.m. <i>Independence Foyer</i>	Networking Break <i>Meet the Editor of APA's NEW journal: Technology, Mind, and Behavior</i>			

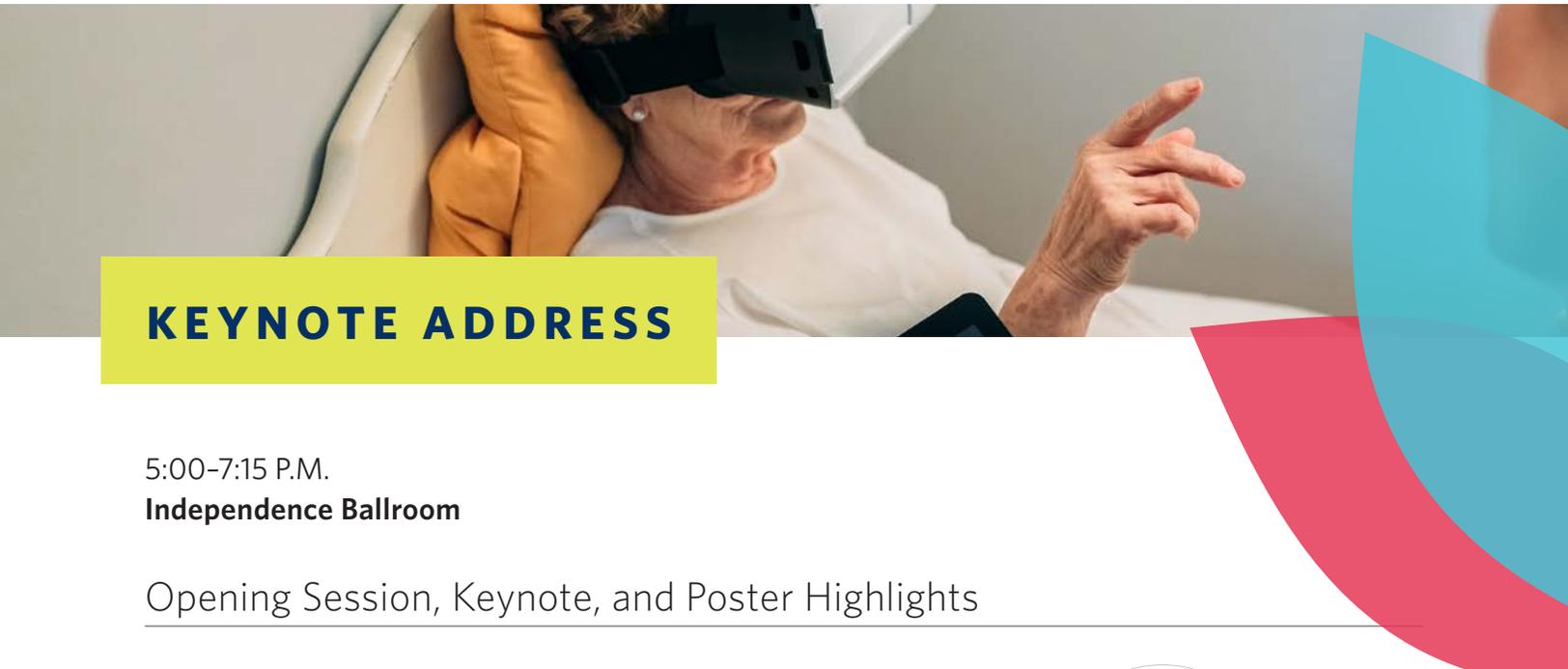
## AGENDA AT A GLANCE

**Earn CE credits at TMS 2019:** Select sessions at this conference have been reviewed and approved by the American Psychological Association's (APA) Office Continuing Education in Psychology (CEP) to offer Continuing Education (CE) credit for psychologists. Full attendance is required at each session for which you are claiming CE credit. Partial credit is not awarded. The CEP Office maintains responsibility for the delivery of the session.

### SATURDAY CONTINUED

<p>11:00 a.m.- 12:15 p.m.</p> <p><b>Concurrent Sessions 4</b></p>	<p>Challenges and Opportunities for Using Big Data <i>Paper Session.</i> <i>Independence D.</i></p>	<p>Social Media and Well-Being <i>Paper Session.</i> <i>Independence E.</i></p>	<p>Measuring Psychological Variables using Mobile Sensing Technologies: Modeling Big Data and Implications for Research and Designing Intelligent Support for Well-Being and Productivity at Work <i>Symposium.</i> <i>Independence FG.</i></p> 	<p>Learning and Education in the Digital Age <i>Paper Session.</i> <i>Farragut Square.</i></p>
	<p>Teletherapy <i>Paper Session.</i> <i>Independence HI.</i></p> 	<p>Dietary Behavior and Technology <i>Paper Session.</i> <i>Independence BC.</i></p> 	<p>Technology in Tandem: Characteristics and Effects of Joint Media Engagement in the Digital Age <i>Symposium.</i> <i>Lafayette Park.</i></p> 	
<p>12:15-1:30 p.m.</p>	<p>Lunch on your own</p>			
<p>1:30-2:45 p.m.</p> <p><b>Keynote</b></p> <p><i>Independence Ballroom</i></p>	<p><b><u>The Digital Revolution: The Potential Promise and Ethical Perils in Research</u></b> <i>Camille Nebeker, EdD, MS</i></p> 			
<p>3:00-4:15 p.m.</p> <p><b>Concurrent Sessions 5</b></p>	<p>Children and Technology <i>Paper Session.</i> <i>Independence BC.</i></p> 	<p>Work and Management in a High-Tech World <i>Paper Session.</i> <i>Farragut Square.</i></p>	<p>Machine Learning and Education <i>Paper Session.</i> <i>Independence E.</i></p>	<p>Engaging Patients and Transforming Care <i>Paper Session.</i> <i>Independence HI.</i></p> 
	<p>The Candid Body-Worn Camera? How Officer Body-Worn Cameras Inform (and Misinform) Our Understanding of Police Encounters <i>Symposium.</i> <i>Independence D.</i></p>	<p>Transforming Access to Mental Health Care for Rural and Underserved Populations <i>Symposium.</i> <i>Independence FG.</i></p> 	<p>Reflections on Cyberlearning: Exploring Tensions in the Co-Evolution of Learning Technologies with Advances in Learning Theories and Methods <i>Panel.</i> <i>Lafayette Park.</i></p>	
<p>4:30-5:30 p.m.</p> <p><i>Independence Ballroom</i></p>	<p>Closing Panel</p>			





## KEYNOTE ADDRESS

5:00–7:15 P.M.

**Independence Ballroom**

Opening Session, Keynote, and Poster Highlights

# Is Clinical Virtual Reality Ready for Primetime?

Since the mid-1990s, a significant scientific literature has evolved regarding the outcomes from the use of what we now refer to as Clinical Virtual Reality (VR). This use of VR simulation technology has produced encouraging results when applied to address cognitive, psychological, motor, and functional impairments across a wide range of clinical health conditions. This presentation addresses the question, “Is Clinical VR Ready for Primetime?”

After a brief description of the various forms of VR technology, Dr. Rizzo will discuss the trajectory of Clinical VR over the last 20 years and summarize the basic assets that VR offers for creating clinical applications. The discussion then addresses the question of readiness in terms of the theoretical basis for Clinical VR assets, the research to date, the pragmatic factors regarding availability, usability, and costs of Clinical VR content/systems. This will be presented in the context of descriptions and video examples of applications addressing Anxiety Disorders, PTSD, Addiction, Depression, PTSD, Pain Management, Stroke, TBI, ADHD, Autism, and Virtual Human applications for clinical training and patient facing healthcare support. Ethical issues for the safe use of VR with clinical populations will then be detailed.

While there is still much research needed to advance the science in this area, Dr. Rizzo will make the case that Clinical VR applications are in fact “ready for primetime” and will soon become indispensable tools in the toolbox of healthcare researchers and practitioners and will grow in relevance and popularity in the near future as the technology continues to evolve.



## **ALBERT “SKIP” RIZZO, PhD**

**Director, Medical Virtual Reality**  
*Institute for Creative Technologies*

**Research Professor**  
*Department of Psychiatry and School of Gerontology, University of Southern California*

5:00–7:15 P.M.

## Independence Ballroom

Opening Session, Keynote,  
and Poster Highlights

---

### Introduction

*Russell Shilling, PhD, Chief Scientific Officer,  
American Psychological Association*

### Opening Remarks

*Arthur C. Evans, Jr., PhD, Chief Executive Officer,  
American Psychological Association*

### Keynote Address

Is Clinical Virtual Reality Ready for Primetime?

*Albert “Skip” Rizzo, PhD, Institute for Creative Technologies  
and University of Southern California*

### Poster Highlights

*David W. Ballard, PsyD, MBA, Senior Director for Applied Psychology,  
American Psychological Association*

7:15–8:30 P.M.

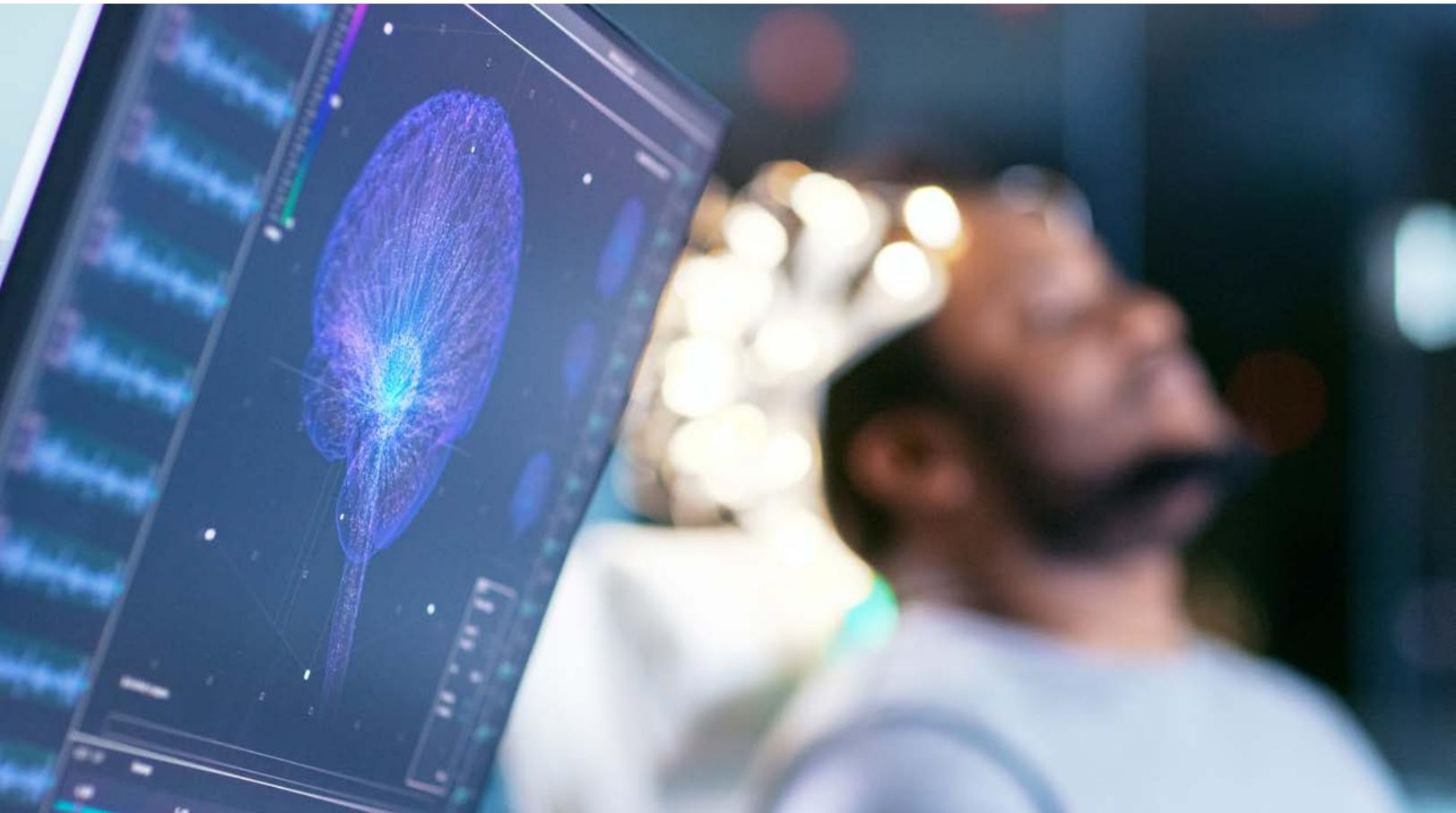
## Constitution Ballroom

Welcome Reception

---

### Welcome

*Rosie Phillips Davis, PhD, President,  
American Psychological Association*





# TECHNOLOGY, MIND & SOCIETY

AN APA CONFERENCE

**OCTOBER 4**

**FRIDAY**



7:30–9:00 A.M.

Constitution Ballroom

## Continental Breakfast / Poster Session 1

### Aging / Older Adults

#### A-1

##### Feasibility of Assessing Cognitive State in the Elderly from Speech Collected via the Telephone

The diagnosis of mild cognitive impairment (MCI) or dementia due to Alzheimer's disease (AD) requires a comprehensive history and clinical examination, which is time-consuming, expensive, and difficult to physically access. There is a pressing need for rapid, inexpensive and easily accessible screening tools in the preclinical and clinical stages of AD, for optimal diagnosis, prognosis, and medical management. Here we report on the feasibility of using natural language processing methods on speech derived via the telephone, in order to develop sensitive metrics that may be used as a potential biomarker. As part of an ongoing study, community-dwelling older individuals who are cognitively unimpaired or diagnosed with MCI or AD participated in an audio-recorded telephone interview. The interview included a standard cognitive screening tool (a modified version of the Telephone Interview for Cognitive Status), and additionally collected speech samples from verbal fluency tasks and in response to a prompt to generate free speech. Elderly volunteers were called at home, the interview was digitally recorded, and participants rated their levels of enjoyment and stress associated with sections of the interview. We address six questions of interest: (1) Will elderly people agree to participate in a recorded telephone interview? (2) Will they complete it? (3) Will they judge it to be an acceptable approach to cognitive screening? (4) Will a sufficient amount of speech data be collected over the telephone for computational methods to be applied? (5) Will the speech data be intelligible to human raters? (6) Will the speech data be of sufficient quality that automated speech recognition can be used for transcription purposes? Preliminary findings include participants' overwhelming agreement to be interviewed, excellent completion rates for the entire interview, and ratings of enjoyment and anxiety for different sections of the interview. These results also demonstrate that sufficient speech data can be collected such that clinically meaningful computational language metrics can be derived. Finally, we will present data on the intelligibility of the speech samples to human raters, and to automated speech recognition software. Our findings demonstrate the feasibility and ecological validity of this approach for the early detection of cognitive decline in the elderly.

*Catherine Diaz-Asper (Marymount University)*

#### A-2

##### Problematic Social Media Use and Perceived Social Isolation in Older Adults

Social isolation exists when individuals lack true social engagement and quality relationships with others<sup>1</sup>. This state, whether real or perceived, is associated with numerous detrimental health conditions and mortality, especially in older adults<sup>2</sup>. In today's digital society, if an individual perceives that they are socially isolated, they can take steps to interact with others online, on social networking sites (SNSs) like Facebook, Twitter, and Instagram. Over 2.5 billion people around the world use SNSs to satisfy their need to connect with others<sup>3,4</sup>, and the prevalence of SNS use in older adults is considerable, with 64% of individuals aged 50–64 years of age and 37% of individuals 65 years and older using these sites in the United States<sup>5</sup>. Importantly, previous research in young adults has demonstrated a relationship between social media use and social isolation; young individuals who report more SNS usage also report more perceived social isolation<sup>6</sup>. With regard to older adults however, not much is known about the interplay of SNS use and social isolation.

Only a handful of studies have investigated the relationship between SNS use and social isolation in older adults, and the results have been mixed<sup>7</sup>. For example, qualitative research has demonstrated that older adults report feeling “less lonely” when using an SNS<sup>8</sup>. Seemingly at odds with this, several quantitative studies have compared older adults who use SNSs to older adults who do not use SNSs, revealing no difference in loneliness between the groups<sup>9–11</sup>. Finally, with a lifespan sample consisting of 15–75 year olds, SNS users were revealed to be lonelier than non-users<sup>12</sup>. Thus, prior quantitative studies on perceived social isolation in older adults have only compared SNS users with SNS non-users. As of yet, no study has investigated perceived social isolation in older adults who use SNSs, examining if perceived social isolation varies with degree of SNS use.

We hypothesized that the positive relationship seen between social isolation and degree of SNS use in younger adults<sup>6</sup> is also manifest in older adults. To address this, we collected online survey data from individuals 50 years of age and older ( $n=214$ ). We obtained demographic information, living situation, and self-reported perceptions of social isolation using the PROMIS Social Isolation Scale<sup>13,14</sup>. Importantly, we also used two measures to ascertain degree of social media use: (1) we asked participants to estimate their daily time on SNSs, and (2) we had participants fill out the Bergen Social Media Addiction Scale<sup>15,16</sup> to assess their degree of problematic, excessive SNS use.

After collecting our data, we conducted a hierarchical logistic regression to better understand the relationship between the factors in our models—we wanted to ascertain the extent that demographic variables, including living situation (Block 1), estimated daily minutes on social media (Block 2), and problematic social media use (Block 3) predicted perceived social isolation in people aged 50 and older. Results are presented in Table 1. Overall, our model was significant

at all three steps, explaining 5% (Block 1), 5% (Block 2) and 13% (Block 3) of the variance in perceived social isolation. Counter to our hypothesis, we did not find that estimated daily minutes on social media, when added to our model, was significantly associated with perceived social isolation. However, in line with our hypothesis, we found that problematic social media use was significantly associated with perceived social isolation. In other words, the more problematic an individual's use of social media is, the more likely they perceive themselves to be socially isolated.

We note that our results are correlational in nature, and causal relationships cannot be determined with the current cross-sectional dataset. Therefore, the association we revealed can be interpreted in one of two ways: (1) the more socially isolated older adults feel, the more they display problematic SNS use, or (2) the more problematic older adults' use of social media is, the more socially isolated they feel. Further research is needed to determine which scenario is correct. Understanding the causal directionality with future research is critical, especially considering the introduction of technological interventions to combat social isolation in older adults (for review see Baker et al.17). If the latter scenario is true, it could be that interventions which introduce social media to counter social isolation in older adults may lead to maladaptive, problematic use of these sites in certain individuals, possibly leading to increased perceptions of social isolation in these individuals. This unfortunate circumstance would yield the exact opposite effect that the intervention's intended goals. With this in mind, our findings provide an important starting point for future studies on problematic social media use and social isolation in older adults (e.g., studies which are longitudinal in design).

*Dar Meshi (Michigan State University)*

### A-3

#### **YouTube Buffers Acute Distress Among a Complex Older Adult Veteran Cohort**

While non-pharmacological management of distress among an older adult veteran cohort with complex co-occurring medical and psychiatric conditions living in long-term care facilities is of critical importance, traditional psychotherapy interventions pose challenges (Bohart & Tallman, 1999, Hartz, Brennan, Jasdeep, & Estrin, 2010, Manepalli, Desai, & Sharma, 2009). To bolster management of distress among the Veteran cohort within a VA extended care Community Living Center (CLC), an array of mobile applications were utilized with veterans as part of their ongoing mental health care treatment plans. Among the applications, YouTube was accessed with Older Adult Veterans to help them to listen to music, access video clips, and review information on Veteran specific interests to facilitate engagement in pleasant events and thereby reduce distress. Clinical impact of engaging in these activities was gathered as part of ongoing clinical program development, process improvement efforts, within the clinical environment. A sample of 30 Veteran's were provided technology supported psychotherapy interventions that included the use of YouTube mobile application. A dependent samples T-Test was utilized to explore if the technology-enhanced intervention had a significant effect on patient self-reported distress. Findings indicated that there was a significant difference between average veteran rated subjective distress ratings before intervention and average veteran rated subjective distress ratings after intervention ( $t(29)=6.618, p<.001$ ). The effect size for the difference in subjective

distress ratings was large ( $d=1.207$ ). Veterans were also evaluated by a provider rated multidimensional behavioral distress scale and there was also a significant difference between average veteran behavioral distress across pre and post intervention ( $t(29)=6.846, p<.001$ ). This difference was of a large effect ( $d=1.250$ ). These results suggest that use of YouTube may help to reduce self-reported and behavioral distress often present among older adult veterans with complex co-occurring medical and psychiatric conditions living in long-term care facilities.

*Peter Louras (Palo Alto University / Palo Alto Veterans Institute for Research)*

## Assistive Technology

### B-1

#### **Development of an Adaptive Skills Training Program for Youths with Developmental Disabilities Using Video Modeling and Pictorial Cues**

Youths with developmental disabilities (DD) show deficits in adaptive skills (Ditterline, Banner, Oakland, & Becton, 2008), which are critical predictors for independence (Matson, Taras, Sevin, Love, & Fridley, 1990), and high level of quality of life (Bathelt, Haan, & Dale, 2019). Adaptive skills are improved with age even for persons with DD, the gap between youths with DD and typically developing children gets wider as time goes on (Bal, Kim, Cheong, & Lord, 2015). Behavioral intervention is proven to be effective to improve adaptive skills of youths with DD (Davis, Smith, & Donahoe, 2002, Eldevik, Jahr, Eikeseth, Hastings, & Hughes, 2010). However, it requires tremendous time and efforts for service delivery, resulting in limited dissemination. With advances of ICT technology, technology-based treatment has become a promising alternative as a service delivery method. The effectiveness of technology-based treatment on adaptive behaviors is supported by previous studies (Artoni, Bastiani, Buzzi, Buzzi, Curzio, Pelagatti, & Senette, 2018; Buzzi, Buzzi, Perrone, & Senette, 2018). In order to identify adaptive skills that are highly demanded in the field, a need assessment and production process were performed. As a need assessment, semi-structured interviews with 5 BCBA's (Board Certified Behavior Analysts), 10 teachers and 17 parents were conducted. Upon the basis of thematic analysis for qualitative data (Braun & Clarke, 2008), contents were classified into six categories; self-help skills (5), situation awareness (8), social rules (2), coping skills (13), communication (5), attention skills (8). Especially, six out of 13 respondents who reported that they expect coping skills to improve through the training program also replied in self-help skills and situation awareness. This data could be used to examine what kind of coping skills are most demanded; utilization of self-help skills and knowing the rules of daily situations. As a result of the thematic analysis, the most demanded training contents were identified as learning situations such as self-help skills and social rules. Although Self-help skills training is effective when it is conducted in an actual workshop setting, or community setting (Cannella-Malone & Schaefer, 2017), video modelling (McCoy & Hermansen, 2007; Shrestha, Anderson, & Moore, 2013) and pictorial cues (Connis, 1979) were also effective as a technology-based intervention. Since there was no previous ICT program available to train basic living skills for youths with DD using video modelling and pictorial cues, a list of self-help skills was created for the development of a new program.

To choose the most basic and important skills for youths with DD to live an independent life, situations were selected by reviewing a book "I can do it alone (analysis of daily life movement)" (Kim, Kwak, Lee, & Kim, 2004), Vineland Adaptive Behavior Scales (Sparrow & Cicchetti, 1989), 11 times of internal meeting, and consultation with ABA specialists. Videos were 15seconds to 1minute long, and pictorial cues were 3 to 5 sequential pictures edited from the videos. Examination of the effectiveness of the program is in progress. Based on the previous researches that Virtual Reality (VR) was an effective rehabilitative intervention for people with intellectual disabilities, and the skills required through VR training were transferred to the real world (Standen & Brown, 2005), the final product of adaptive skills training program is expected to be utilized to produce Virtual Reality (VR) program for youths with DD.

*Dongjoo Chin (Yonsei University)*

## Autonomy / Autonomous Vehicles

### C-1

#### **Designing Displays in Automated Vehicles for the Thrill-Seeking Driver: Characterizing Driving Profiles to Create Personalized Driving Displays**

Automation, or the performance of a task by a machine or system that might otherwise be completed by a human, surrounds us in many everyday tasks, from autocorrecting of spelling to automated vehicles (Parasuraman & Riley, 1997). Automated vehicles can range in level of automation from highly automated (SAE Level 5), requiring no input from the driver, to low automation (SAE Level 1) such as blind spot warning systems (SAE International, 2013). As automation levels increase, these systems drastically change the driving task, ultimately relegating the driver into a supervisory role in the control of the vehicle (Sheridan, 2012)

However, automation, including in vehicles, is not always perfectly reliable. In fact, in complex environments, such as on the roadways, automation can have quite variable reliability and can even fail (Dikmen & Burns, 2016). Providing the driver with a display of the automation's level of (un)certainly in its own performance has been shown to improve the experience of driving an automated vehicle, even after experiencing an automation failure, by supporting trust calibration (Noah, 2018). The research we report on, here, builds upon that automation uncertainty display research, with the end goal of improving acceptance and trust of highly automated vehicle technologies (Noah & Walker, 2018).

Having the vehicle automation system present information to the driver about what the automation is doing, the choices the system is making, and the objects detected around the vehicle aids transparency and therefore trust calibration (Hoff & Bashir, 2015; Noah & Walker, 2017). However, different drivers may want (or need) different information or the information presented in different ways. Having a variety of display approaches, or profiles, personalized to the wants and needs of various driver types is certainly possible. The question then arises how drivers characterize these different display profiles, what driving behaviors are associated with them, and what display elements would support specific profiles. The driving profiles explored in this particular study were: Thrill-Seeking, for the driver who finds enjoyment in driving, and wants information about speed,

and vehicle performance; Transit, for the driver who wants to know the general route, and perhaps fuel efficiency; and Defensive, which provides information about what the vehicle detects in the environment, for the driver concerned with safety (Noah & Walker, 2018). These displays could even be used to gradually gain the trust of a driver who is new to highly automated vehicles. For example, the defensive display could be used initially to build the driver's trust, and then transition to the transit display. For someone who is a driving hobbyist, the thrill-seeking display may be what entices them to use the automation systems.

To support our research goals, we enlisted 25 undergraduate students from a large US research university to participate in focus groups that discussed the similarities and differences of each driving profile, and developed display prototypes (for the thrill-seeking profile, in this particular study). All participants were required to have corrected or normal-to-corrected vision, corrected or normal-to-corrected hearing, and be 18 years or older or enrolled as a student. One female moderator facilitated each of the focus group sessions, while an additional researcher took written notes.

The moderator followed a discussion guide that included the following components: (1) a description of the purpose for the study and rules for the focus group; (2) an outline of the topics and corresponding discussion questions; and (3) instructions for the design activity. A fixed sequence of 13 discussion questions was used to explore three topics (knowledge of automated vehicles, identification of driving profiles, and perceptions of thrill-seeking driving behavior). Table 1 includes all topics and discussion questions used to guide the focus groups. To conclude the session, participants completed a design activity consisting of four components: initial design prototyping; narrowing down design ideas; feedback; and a final group design (Table 2). The goal of the design activity was to gain insight on what information and design elements should be included in displays for the thrill-seeking profile.

Affinity diagramming is being used now to identify themes of thrill-seeking driving behavior and common elements among the displays from the design activity. The information gathered from this study will guide our understanding of driving profiles and the expectations of drivers. The results will then be used to improve display designs so they match the mental models of future users. Utilizing personalized driving displays in automated vehicles can help the driver understand how the automation is performing. By improving the driver's situation awareness, they can be more prepared to transition from automated driving back to manual driving. In addition, using displays that match the driver's personal driving profile may improve trust in automation and ensure the driver has an appropriate trust calibration. Future research will investigate how the use of personalized driving displays impacts situation awareness and trust in automated vehicles.

*Zoe Becerra (Georgia Institute of Technology)*

### C-2

#### **Perceptions of Terms Used to Describe Automation in Vehicles: A Multidimensional Scaling Study**

Concerns have arisen about the vocabulary that has been used to describe automation in vehicles. In particular, Tesla has been criticized for using terms like "Autopilot" and "full self-driving," because these terms seem to imply functionality beyond the current

capabilities of the systems they describe. Research has suggested that perceptions of the functionality implied by some of the terms used to describe automation in vehicles do not align with the actual capabilities of automated systems (see Abraham, Seppelt, Mehler, & Reimer, 2017; Nees, 2018). This exploratory study examined perceptions of a sample of terms used to describe automation in vehicles using similarity ratings analyzed with multidimensional scaling (MDS).

**Method: Participants.** Participants (N = 303) were recruited online via Amazon Mechanical Turk. Twenty-six participants failed attention or eligibility checks, so the final sample analyzed was N = 277 (130 females, 145 males, and 2 people who chose “other/prefer not to respond”; M age = 34.69 years, SD = 10.34). Survey. Participants rated the similarity of pairs of terms that have been used to describe automation in vehicles. Figure 1 shows the 39 different terms that were examined. For each pair, participants rated the extent to which they perceived that the terms indicated similar functional capabilities of the vehicle. Terms that could be used as an adjective to modify “vehicle” were presented as such; all other terms were presented as the object of the phrase “a vehicle with”. Procedure. For each participant, a Qualtrics survey presented 50 of the possible pairings of the terms. Pairs were pseudo-randomly selected to approximately balance the overall number of times a given pair of terms was presented across all participants. Across all participants each pair of terms was rated M = 18.69 (SD = 4.17) times. The order of the appearance of the terms in the question (i.e., first or second) was randomly determined for each pair of terms for each participant.

**Results and Discussion:** Ratings were averaged across participants and across order of presentation to form a symmetric matrix of aggregate similarity ratings. The averaged similarity ratings (scaled from 1-7) were subtracted from 8 to re-scale the ratings as dissimilarities. The matrix of dissimilarity ratings was modeled with classical MDS using the non-metric (ordinal) ALSCAL procedure in SPSS. The model with two dimensions (see Figure 1) was selected as the most appropriate model for interpretation; the model produced a stress value of .26 and R2 = .80. The first dimension clearly corresponded to the perceived level of automation, with terms toward the left side of the space corresponding to lower levels of automation.

In some instances, the model suggested that participants’ perceptions of the terms did not align with the capabilities associated with the technical definitions of the terms. “Autopilot” was rated as relatively similar to “autonomous” and “self-driving.” For other terms, participants underestimated the capabilities of the vehicle according to the technical definitions. “Unsupervised driving mode,” for example, is described by Volvo to mean “the vehicle does all the driving,” yet this term was perceived to be similar to “function-specific automation” and “driver assistance.” Further, “unsupervised driving mode” was rated as similar to—but entailing a slightly lower level of automation than—“supervised driving mode.” Participants apparently were unclear on whether the term implicated the driver or the vehicle as the supervisor. “Traffic Jam Pilot,”—the term for Audi’s system, was perceived to entail lower levels of automation than traditional “cruise control” and was rated most similar to “lane trace control” and “lane centering steering” on Dimension 1, which suggested that the term is perceived to imply less functionality than the actual capabilities of the system.

Dimension 2 was more difficult to interpret, but the anchors suggested that Dimension 2 might represent functions that apply to braking (with “autobrake” as one anchor) versus accelerating (with

“cruise control” toward the other anchor). This raises an important qualification of the interpretation of this model, namely that the MDS procedure is exploratory, and the stress and R2 values indicate that the model obtained offered a noisy and incomplete description of perceptions of terms related to automation. This is not surprising given the presumably high degree of variability in participants’ impressions of what the terms mean.

*Michael Nees (Lafayette College)*

### C-3

#### The Effects of Vehicle Automation Level and Warning Type on Responses to Vehicle Hacking

Modern surface transportation involves vehicles with different levels of automation (SAE, 2018). Automated and connected vehicles with higher levels of automation are at a higher risk for hacking attempts because automated driving assistance systems often rely on onboard sensors and internet connectivity (Amoozadeh et al., 2015). Although research has examined the effect of level of automation on driver situation awareness (SA; Endsley, 1996) and behaviors following vehicle control take-over (Gold, Damböck, Lorenz, & Bengler, 2013; Merat, Jamson, Lai, & Carsten, 2012; Shen & Neyens, 2017; Strand, Nilsson, Karlsson, & Nilsson, 2014), it is unclear how humans develop SA and respond to novel events like a vehicle hacking attempt. Specifically, the current study examines the impact of vehicle automation on drivers’ ability to develop SA and respond to a vehicle hacking attempt, using a medium-fidelity driving simulator.

Auditory warnings can effectively attract a driver’s attention while performing a driving task which is often visually demanding (Baldwin, 2011; Petermeijer, Doubek, & de Winter, 2017). An auditory warning can be either speech-based containing semantic information (e.g., “car in blind spot”) or non-semantic (e.g., a tone, or an earcon), which can influence driver behaviors differentially (Sabic, Mishler, Chen, & Hu, 2017). For example, McKeown and Isherwood (2007) showed that, non-semantic, tone-based warnings impeded performance while speech and auditory icon warnings improved performance in a car warning recognition task. The semantic content contained in speech warnings was also found to effectively reduce drivers’ reaction times better than tones when the situation was more complex (e.g., in an intersection) (Chang, Lin, Fung, Hwang, & Doong, 2008). Based on these previous findings, drivers benefit from auditory warnings containing semantic information depending on the complexity of situation. Responding to a hacking attempt is a complex situation that is time sensitive, so the authors plan to examine the effect of semantic warnings on driver responses to hacking attempts.

**Research problem.** The purpose of the current study was to examine the effect of level of automation and warning type on driver responses to novel critical events, using vehicle hacking attempts, in a driving simulator. A better understanding of how humans respond to vehicle hacking attempts as well as factors that affect their responses, will help promote the development of effective vehicle automation in the future.

**Procedure.** The current driving simulator experiment manipulates two independent variables. Level of Automation (Manual vs. Automated) and Warning Type (Semantic vs. Non-Semantic). A full factorial design with the two factors made four distinct conditions (Manual Semantic, Manual Non-Semantic, Automated Semantic,

and Automated Non-Semantic). Participants completed two simulated drives in a STISIM 3 driving simulator where they were asked to drive to school. The first drive ended with the participant safely arriving at school. During the second drive, a two-second warning was presented to the participant three quarters of the way, immediately followed by a simulated vehicle hacking attempt. The warning either stated "Danger, hacking attempt incoming" in the semantic condition or was a 500 Hz sine tone in the non-semantic condition. The vehicle hacking attempt caused the vehicle to swerve off of the road and towards buildings and parked cars. This event lasted exactly 5 seconds before simulating a crash into a vehicle and ending the simulation if no intervention by the driver occurred. The dependent variables collected to measure driver responses to the hacking attempt included: Time to Collision (TTC), Driver Response Method (wheel turn/pedal press), and Response Rate (the rate of successful response to the hack).

Planned data analysis. Data collection is currently ongoing. A 2 (Level of Automation) x 2 (Warning Type) factorial ANOVAs will be performed to examine the differences between the different experimental conditions for TTC and Response Rate. A chi-squared analysis will be used to examine if there are different methods of response to the hacking attempt among the experimental groups.

Expected results and contribution. The current study aims to present evidence that automation level and auditory warning type affect driver responses to vehicle hacking attempts. Based on past empirical findings and Endsley's (1996) model of SA we expect that the Manual Semantic and Manual Non-Semantic groups will have higher TTC values, and better response rates than participants in the Automated Semantic and Automated Non-Semantic groups. We also expect that the Manual Semantic and Automated Semantic groups will have higher TTC values, and better response rates than participants than their respective partner groups (Manual Non-Semantic and Automated Non-Semantic). Based on these expected findings, manual driving and semantic warnings should be used when the driver is required to respond to a vehicle hacking event. Future vehicle designers may consider the results of the current study for the design of driving assistance systems using more effective warning types.

Scott Mishler (Old Dominion University)

## C-4

### "Well, That Wasn't So Bad!": Comparing Forecasted and Retrospective Ratings of Affect in the Context of Automated Driving

Affect is a transient emotional state that has both a positive and negative valence (Watson, Clark, & Tellegen, 1988). Positive affect (PA) is related to things such as excitement and arousal (Watson et al., 1988). Negative affect (NA) is related to things such as fear, sadness, and frustration (Watson et al., 1988). The Positive and Negative Affect Schedule (PANAS) was developed to measure self-reported positive and negative affective states in the present or past (Watson et al., 1988). PANAS has been adapted for forecasting affect, or estimating how one would likely feel in relation to an anticipated situation (Calderwood, Green, Joy-Gaba, & Moloney, 2016; Noah, Gable, Schuett, & Walker, 2016). People tend to overestimate the NA that they expect to experience in a given situation, when compared to their reported affect during the situation (Calderwood et al., 2016). Forecasting affect has also been used to determine how different driving environments (highway, rural, urban, and suburban) could

impact feelings towards automated safety features and driving systems (Noah et al., 2016).

The goal of the present study was to determine how forecasted PA and NA compared to post-experience affect, in the automated driving context. Forecasted affect is particularly important in this context, as it could be a barrier to acceptance of automated vehicles. If people anticipate negative feelings from driving an automated vehicle, they may choose to avoid that technology.

Here, 62 participants completed a forecasting PANAS prior to interacting with an automated vehicle simulation. They were to imagine driving a vehicle and using an automated lane keeping system (ALK; a system that controls lateral lane position), when responding to the items. Participants then drove a 7-8 minute simulated drive (baseline drive) using ALK, after which they completed a (retrospective) PANAS assessing the drive they just completed. Participants then completed an approximately 15-minute drive (failure drive), also with ALK. During this drive, there was an automation failure (ALK system turned off), from which participants had to recover and continue to drive manually. Participants completed the last PANAS after the failure drive, with instructions to refer to the drive they just completed when responding to the items.

We hypothesized that PANAS timing (forecasted, after baseline drive, and after failure drive) would impact both positive and negative affect scores. Specifically, we hypothesized that there would be a significant change in both PA and NA scores, in reference to the forecasted and the post-baseline drive and post-failure drive measurements.

Results from this study, seen in Figure 1, showed that there were significant main effects of timing (forecasted, baseline drive, failure drive) for both PA,  $F(1.486, 4.459) = 8.997, p = .001, \eta^2 = 0.134$ , and NA,  $F(1.742, 5.525) = 37.554, p < .001, \eta^2 = .393$ . Follow-up paired t-tests for both PA and NA used the Bonferroni corrected alpha of  $p = .0167$ . There was no significant difference between the forecasted PA and the baseline drive PA, which indicates initially experiencing a neutral (normally functioning) automated driving system was congruent with participants' expectations. After experiencing an automation failure in the failure drive, participants' PA was significantly reduced, in comparison to the forecasted PA,  $t(61) = 3.451, p = .001$ . There was also a reduction in PA between the baseline and failure drives,  $t(61) = 2.583, p = .012$ . This shows that even after an initial neutral interaction with the automation, the automation failure greatly impacted PA.

There was a different, but also interesting, pattern of results for the NA. The forecasted NA was significantly greater (i.e., more negative) than what was reported after the baseline drive,  $t(61) = 5.658, p < .001$ . That is, participants felt less negative about the automation after they experienced it. Most interestingly, there was a further reduction in NA between the baseline and failure drives,  $t(61) = 3.435, p = 0.001$  (and also between the forecasted and failure drive scores,  $t(61) = 7.243, p < .001$ ). Even though participants had a negative experience with the automation, their resulting level of NA was nevertheless diminished. The root of this result is a substantial over-estimation of the forecasted NA. After the first (neutral) experience, participants seem to have realized automated driving was not as bad as initially expected. These negative feelings were further reduced by experiencing the automation failure. This pattern of results is similar to previous research where students overestimated their NA with multimedia sources being removed from a homework task (Calderwood et al., 2016).

The results of this extend previous results to the automated driving domain showing that, fundamentally, people tend to overestimate the negative feelings they will have with a technology or situation prior to experiencing it. These results highlight the importance of experiences with automated vehicles in shaping our feelings towards them. Future work should explore how initially negative interactions may impact feelings about automated systems over time.

*Brittany Holthausen (Georgia Institute of Technology)*

## Avatars / Agents

### D-1

#### Virtual Humans in Clinical Interviewing: The Role of Visual Anonymity

Research on computer-mediated communication has demonstrated that computer-based assessments can reduce barriers to self-disclosure. This empirical evidence has recently been applied in the context of clinical interviewing through virtual humans. Lucas et al., (2014) showed that participants who believed to be interacting with an agent (an autonomous virtual human controlled by a computer program) self-disclosed more than participants who believed to be interacting with an avatar (a virtual human tele-operated by another person) during a clinical interview. The results suggested that the perceived agency of an interaction partner (human vs. computer) could have an effect on self-disclosure. However, both conditions also differed in their perceived visual anonymity, which may have also contributed to these findings. Whereas in the avatar condition participants were led to believe that they would be visible to their interviewer, in the agent condition participants were led to believe that they would have visual anonymity, as only their audio recordings would be accessed by researchers. The current study aimed to disentangle the effects of agency and visual anonymity on self-disclosure by introducing a third condition—an avatar tele-operated by a person without access to participants' video feeds. College students (N=88) were randomized to three conditions in which they were led to believe that they would either interact with an agent, an avatar with video access, or an avatar without video access. All participants interacted with a computer program to ensure consistency. Following their interaction with the virtual human, participants reported on the depth of their self-disclosure during the interview. Results of the omnibus test using the Brunner-Dette-Munk method revealed significant differences in self-disclosure across conditions ( $F = 3.43, p = 0.03$ ). Pairwise comparisons demonstrated that individuals reported greater depth of self-disclosure in the avatar without video access condition than in the avatar with video access ( $t(33.69) = 2.28, p = 0.03$ ) and agent condition ( $t(35.89) = 2.80, p = 0.01$ ). The latter two conditions did not differ ( $t(35.21) = 0.78, p = 0.44$ ), failing to replicate the results by Lucas et al. (2014). Our results suggest that lack of visibility during an interaction increases self-disclosure, which might be due to individuals experiencing more feelings of anonymity and less fear of negative evaluation. When visual anonymity was held constant across avatar and agent conditions, significant differences emerged with participants self-disclosing more towards a human. It is possible that participants perceived higher rapport with a human than with the computer and were therefore more willing to self-disclose. Future

research is needed to elucidate the mechanisms by which technology increases self-disclosure and its applications in clinical settings.

*Laura Garcia (University of Southern California)*

## Big Data / Data Mining

### E-1

#### Adolescent Depression Screening & Clinical Data Use in the ModEHRn Age

Statement of the Problem: Large scale health organizations have been instrumental in the adoption of electronic health records (EHR). EHRs have allowed easier access to "big data" by quickly aggregating patient level data that is systematically collected at healthcare appointments. However, relying on standard clinical documentation can often lead to inaccuracies in the data and can even lead to significant inferential errors (Kaplan, Chambers, & Glasgow, 2014; Madden et al., 2016). This study analyzed data from the recent implementation of universal depression screening for adolescents in a large-scale urban pediatric primary care residency training clinic (Suh et al., 2018), as recommended by the American Academy of Pediatrics (Zuckerbrot et al., 2018), to identify the frequency of errors inherent in the data when relying solely on informatics reporting or electronic data pulls of clinical documentation in EHRs. Procedures: In 2017, universal screening for symptoms of depression was instituted clinic-wide using the Patient Health Questionnaire—A (PHQ-A; Johnson et al., 2002) at all visits for youth between the ages of 11-18. A clinical workflow was developed within the primary care clinic to provide a pathway for patient screening, provider awareness of results, and utilization of support staff to document the depression screening results in the EHR. A formal partnership was then established between the primary care clinic's integrated mental health team and research informatics to be able to aggregate and pull the clinic's PHQ-A data for the purposes of program evaluation and quality improvement. Manual record review and data integrity checks were subsequently conducted to document accuracy of clinical data stored in the EHR. Analyses: Frequencies of both visit-centric and patient-centric data were analyzed to gain better understanding of adherence to the universal depression screening workflow and to determine if clinical indicators of depression severity were being entered, documented and stored correctly in the EHR. A clinical cutoff score of  $\geq 10$  was used on the PHQ-A to classify screens as elevated. Endorsement of item number 9 on the PHQ-A (i.e. suicidal thoughts) was captured and reviewed to gain insight into symptom severity. An elevated score or a positive response on item number 9 (i.e., endorsement of any suicidal thoughts) triggered consultation with integrated behavioral health providers as part of the workflow and response protocol. Results: From 1/3/17-8/31/18, 2,673 adolescent patients (52% male, 26.5% Caucasian) attended 6,095 visits ( $\bar{x} = 2.4$ ) where they were eligible to receive a PHQ-A. Retrospective data pulls of clinical data entered into the EHR, accessed by informatics, identified a depression screening rate of 14.9% (906 visits), whereas manual record review of all 6,095 patient visits yielded a depression screening rate of 60.9% ( $n=3,695$ ). Moreover, solely relying on data accessible via informatics only identified 66 visits (1.8%) with elevated depression screenings, while systematic medical record review yielded 326 visits (8.8%) with elevated depression screenings. Additionally, analysis of

item level depression screening data showed that data accessible only via clinical informatics identified positive endorsement of suicidal ideation at 15 visits (0.4%), when it was actually endorsed at 59 visits (1.6%) identified through manual record review. Furthermore, visit data were consolidated longitudinally by patient (N=2,673), and informatics identified 58 patients (2.2%) with elevated depression screening scores in this time period, whereas manual record review confirmed there were 230 patients (8.6%) with clinically significant PHQ-A scores. During the study time period, 174 patients (6.5%) with an elevated PHQ-A score that were not identified using clinical informatics strategies. Similarly, 113 patients (4.2%) with symptoms of suicidal ideation were not accurately documented in the EHR in a way that allowed for accurate identification using clinical informatics. Practical Implications: Recent guidelines put forth by the American Academy of Pediatrics call for considerable practice transformation to provide universal screening of depression symptoms to adolescent patients. Moreover, clinical workflows must be reviewed in conjunction with informatics processes to ensure accuracy and integrity of clinical data stored and aggregated from the EHR. Conclusions: "Big data" are only useful when accurately and reliably capturing clinically significant aspects of a patient visit. Substantial data cleaning efforts and manual record review may be necessary before clinical data entered into an EHR can be considered stable, reliable, and useful in employing patient care algorithms. Consideration of the patient > provider > EHR clinical workflow is essential when implementing universal screening initiatives, and working collaboratively across departments (e.g. pediatrics, child and adolescent psychiatry, and informatics) can allow for necessary program evaluation and quality improvement efforts to ensure the validity of clinical data.

Ryan Asherin (University of Colorado Denver)

## E-2

### Internet Searches for Medical Symptoms Preceding the Seeking of Twelve-Step Addiction Treatment Information: A Web Search Log Analysis

Background: Brief intervention (BI) is a critical method for identifying patients with addiction in primary care settings and motivating them to consider treatment options. Yet despite considerable evidence of delay discounting in addiction patients (i.e. discounting higher time-delayed rewards in preference for lower immediate rewards), most brief physician advice focuses on long-term negative medical consequences that may not best contribute to their motivation to seek treatment information.

Objective: Identifying which specific symptoms most motivate individuals to seek treatment information may offer insights for further improving brief interventions. To this end, we use anonymized Internet search engine data to investigate which medical conditions and symptoms preceded searches for twelve-step meeting locators and twelve-step general information.

Methods: We extracted all queries made by people in the USA to the Bing search engine from November 2016 to July 2017. These queries were filtered to those which mentioned seeking Alcoholics Anonymous (AA) or Narcotics Anonymous (NA) information, as well as queries which contained a medical symptom or condition or a synonym thereof. We identified medical symptoms and conditions which predicted searches for seeking of treatment at different time lags. Specifically, symptom queries were first determined significantly predictive of subsequent twelve-step queries if the probability of

querying a medical symptom by those who later sought twelve-step information exceeded the probability of that same query being made by a comparison group of all other Bing users in the USA. Second, we examined symptom queries preceding twelve-step queries at 7, 14, and 30-day time lags, where significance was determined if p-value was less than .05 after Bonferroni correction and the probability of asking about a medical symptom was greater in the 30-day time window than all previous times that symptom was queried.

Results: In our 11,784 person sample, we found 10 medical symptoms that predicted AA information-seeking and 9 symptoms that predicted NA information-seeking. Of these symptoms, a substantial number could be categorized as non-severe in nature. Moreover, when medical symptom persistence was examined across a one-month time period, a substantial number of non-severe yet persistent symptoms were also identified.

Discussion: Our findings raise questions about the possible underlying mechanisms that might put otherwise common or non-severe medical symptoms on similar footing as more dangerous medical conditions for persuading twelve-step information-seeking. One possible underlying mechanism is that more common symptoms may better allow individuals to envision the negative medical consequences of Alcohol Use Disorders (AUD's) and Substance Use Disorders (SUD's) compared to more chronic or dangerous symptoms. In other words, since a symptom like heartburn is common, it may be easy to identify with the experience of the symptom, notice it worsening with increasing substance use, and envision it becoming even more persistent or chronic. In contrast, it may be difficult for an individual to envision getting a cancer diagnosis or its likelihood may be discounted as too distant for concern; moreover, for those who can envision a cancer diagnosis, such envisioning may be so frightening that it fosters denial rather than motivation. Emphasizing the potential for worsening frequency and discomfort of otherwise common symptoms (e.g. heartburn, dizziness, hives, back ache) in addition to severe medical consequences could help further motivate individuals to accept treatment after being newly identified in primary care settings as having an AUD and/or SUD. A second possible mechanism underlying our findings may be related to the social embarrassment associated with certain medical symptoms. For example, symptoms such as bloating, bulging eyes, and impotence may impact physical attractiveness. Social conformity has been shown to be an important motive for alcohol and drug use, with alcohol in particular being considered a social lubricant. Adding to BI's current practices, by highlighting how the medical consequences of substance use can be socially embarrassing may also help motivate patients newly diagnosed with alcohol use disorder or substance use disorder to consider treatment options including 12-step programs. Although discussion of the social repercussions of drinking alcohol is already a component of the brief motivational interviewing that is sometimes incorporated into BI, our results suggest it may provide additional motivation to discuss how medical and nonmedical consequences can interact to intensify social embarrassment in patients with newly identified AUD or SUD.

Conclusions: Our results suggest that many common and/or non-severe medical symptoms and conditions similarly motivate subsequent engagement with AA and NA programs. In addition to highlighting severe long-term consequences, brief interventions could also be restructured to highlight common medical symptoms that could worsen in the short-term with increasing substance misuse and how these worsening symptoms could exacerbate the negative social consequences of substance misuse.

George Nitzburg (Columbia University Teachers College)

**E-3****When Big Data is Too Big:****A Focus Group Study to Guide Design of a Brief Web-Based Patient Treatment History Using Medicaid Billing Data**

Background: Big Data presents an opportunity to integrate massive volumes of clinical history data which can provide critical information for treatment providers and has the potential to improve quality of health care. In order to be useful at the point of care, information in clinical decision-making tools need to be rapidly scanned within a fast-paced clinical treatment environment. In 2003, the Office of Mental Health from a large mid-Atlantic state developed a HIPAA compliant, web-based free-ware platform for supporting clinical decision-making and quality improvement initiatives. The platform leverages over 1 terabyte of data to display up to 5 years of information for over 8 million Medicaid enrollees. The clinical summary report integrates information from all treatment settings, including mental health, substance use, and medical inpatient, outpatient, and pharmacy services for a diverse group of over 25,000 health professional users. A series of unpublished key informant interviews with super users highlighted a significant challenge: although clinicians value and rely on the clinical summary for gathering a patient's history, users complained that the volume of data displayed especially for patients with complex medical needs had hindered the utility of the tool. To address this need, the team conducted a qualitative study with key users to guide the design of a brief clinical summary that condenses 5 years of clinical data into a single page highlighting the most clinically meaningful data.

Methods: Similar to the approach used by Perrier, Kealey, & Straus (2014), the team conducted 18 online or in person focus groups with 348 users across three types of user groups: frequent "super users" (n=92), moderate users (n=30), and new "naïve" users (n=226). Based on the analysis of user logs, "super users" reflected users who viewed over 1,000 clinical summaries during the third quarter of 2017, "moderate users" had seen at least one clinical summary, and "naïve users" were individuals in the process of obtaining access for a quality improvement project. Participants represented staff from 17 hospitals and 18 outpatient mental health clinics. Focus groups were conducted in progressive stages starting with super users and ending with naïve users allowing each subsequent group to provide feedback on iterative designs. Using deductive content analytic methods, team members synthesized participant feedback following each session to examine themes related to user needs and preferences and to guide development of final application design.

Results: All user types including super, moderate and naïve users from both inpatient and outpatient settings valued the concept of a one-page clinical summary that allowed them to drill-in and obtain additional information if they had time and need. Focus groups yielded specific recommendations about content, layout, labeling, sorting, and time period for querying patient data to improve rapid comprehension and ease of use. Although some types of users (e.g., hospital users vs. outpatient users and physician vs. non-physician users) had different preferences for how to sequence data and label clinical information, the order became less important when all information was displayed on a single page. Specifically, the focus groups highlighted four important themes to enhance use: 1) Design should account for users who access the clinical summary online or as a paper printout (i.e., support staff print out the clinical summary for use by a clinician). 2) The way in which data is sorted is important

for comprehension. For example, for patients with multiple diagnoses, users preferred seeing the diagnoses sorted chronologically vs. seeing the diagnoses that were treated most frequently first. 3) Although it is more consistent to show information for a single timeframe, the type of data dictated the most meaningful look-pad time period, for example, users preferred one year for medication history but five years for emergency room and hospitalizations. 4) Data labels/ headings impacted user comprehension and varied by user type; therefore, it cannot be assumed that even common clinical abbreviation would be understood by all users.

Conclusions: Focus group participants viewed a secure, HIPAA-compliant web-based platform for viewing patients' Medicaid service use history as a valuable clinical decision-making tool. Given the diverse user settings and formats, design and functionally that accounts for electronic and printout display, addresses multiple user preferences for sorting clinical data, allows for data queries for different periods of time, and has clear language free of acronyms increases the clinical utility and ease of use for outpatient and inpatient health providers. Future work will involve developing and implementing the brief clinical summary.

*Kristen McLaughlin (New York Office of Mental Health)*

## Cognition / Cognitive Processes

**F-1****Brain or Computer? People Adopt Problem-Solving Strategies That Match Their Performance Goals.**

Most research on human cognition has focused on processes "inside the box". However, without technological support, many humans struggle to solve cognitive tasks like arithmetic (Osiurak, Navarro, Reynaud, & Thomas, 2018; Walsh & Anderson, 2009), spatial navigation (Drive 1 in: Fenech, Drews, & Bakdash, 2010), or prospective memory (Cherkaoui & Gilbert, 2017; Gilbert, 2015) tasks. Thus, to understand how humans solve such everyday cognitive problems, it helps to expand cognitive science's classic focus of what's happening inside the brain and include body (i.e., Embodied Cognition; e.g. Wilson, 2002) and environment (i.e., Situated Cognition; e.g. Kirsh, 2009; Extended Cognition; e.g. Clark & Chalmers, 1998; or Distributed Cognition; e.g. Hollan, Hutchins, & Kirsh, 2000) into the scientific spotlight. For processing that is distributed between brain-based and environment-based resources to work efficiently, humans need to rely on monitoring mechanisms that help them decide when to use which resource. Here, a novel human-computer-interaction paradigm (for Task Description, see Fig. 1) was employed to explore whether people are able to efficiently monitor and mix brain-based and environment-based resources in a goal-directed manner. Specifically, we asked 200 participants to solve a cognitive task, provided them with different performance goals—maximizing speed or accuracy, respectively—and measured how frequently (Experiment 1) and how proficiently (Experiments 2A and B) they made use of a novel HCI device that they could use to support their cognitive processing. Results of Experiment 1 show that participants with different performance goals differed in how frequently they used the novel HCI device. Specifically, participants recruited the device less frequently when aiming for speed (device was used in 80% of trials) than when aiming for accuracy (device was used in 96% of trials).

These results suggest that performance goals influence how we use novel HCI devices. In Experiments 2A and 2B, the focus was on investigating whether the device use exhibited in Experiment 1 is associated with high goal-related performance, i.e. whether the differential device use frequency can be considered adaptive. In both experiments, Experiment 2A and B, we asked one group of participants to exclusively rely on their brain-based resources without availability of the device to solve the cognitive task (forced internal condition) and another group of participants to exclusively rely on the device (forced external condition). We then compared performances in the forced internal and forced external conditions to the performance in Experiment 1, in which participants were able to freely choose between internal and external processing (free choice condition). Results of Experiment 2 show that participants in the free choice condition performed as well as the forced internal group in the speed condition and as well as the forced external group in the accuracy condition (Fig. 2). These results support the notion of the human as a canny offloader who is able to employ cognitive strategies involving external resources in a flexible way that helps fulfill current cognitive goals. Further investigating the specifics and limitations of such a proficiency seems important in an increasingly computerized world that supplies an abundance of external resources.

*Eva Wiese (George Mason University)*

## F-2

### Near-hand Space Effects on Categorization Tasks Resulting from Touchscreen Interface Use

New modes of human computer interaction are shifting from keyboard and mouse-driven interfaces to touchscreen and gesture capture interfaces. One result of this interface change is that an individual's hands may now be in close proximity to the visual information presented on computer displays or in augmented or virtual reality environments. Basic research in cognitive and vision sciences demonstrates that visual processing is altered for objects appearing in the space immediately surrounding the hands, termed near-hand space (NHS; Brockmole, Davoli, Abrams, & Witt, 2013; Goodhew, Edwards, Ferber, & Pratt, 2015). NHS effects occur in basic visual perception processes such as figure-ground segmentation, visual attention, visual working memory, and cognitive control tasks. The current research considers whether NHS effects extend to 1) complex, real-world visual stimuli and 2) categorization and category learning tasks, such as might be important in work environments requiring image analysis.

**Method.** Sixteen participants completed the experiment. Two participants failing to reach criterion on at least one task were dropped from analysis. Stimuli consisted of photorealistic microorganisms modified to have 3 distinct internal features (drawn from Blair, Watson, Meier, 2009; Figure 1). The stimuli subtended 20 cm on-screen display size with a distance of 12 cm between critical features, measured from center to center. Participants sat approximately 57 cm in front of a 24-inch touchscreen monitor (Dell P2418HT). A pair of response labels for the categorization tasks appeared to the right of the target stimulus, positioned 6 cm from the near feature position and 18 cm from the far feature position. Participants received a cover story instructing them to learn the characteristics of a new strain of bacteria (e.g., "Airborne", "Not airborne"). They performed two successive category learning blocks, where the category

structure was defined by one of the internal features, while the other two features were irrelevant. Responses were made via either touchscreen or mouse interaction. The hand was blocked from view by a raised platform during mouse responding. Feedback appeared on every trial. Participants completed at least 40 trials per block, or until the criterion of 10 correct consecutive trials was met. The block terminated after 100 trials if the criterion was not met. Eye movements were monitored using a Pupil Lab head-mounted eyetracker.

The design of the study was 2(Interaction Mode: touchscreen vs. mouse; within) x 2(Critical Feature Position: near vs. far; within) x 2(Frequency of Critical Feature in NHS: 70% vs 30%; between) with Interaction Mode counterbalanced across category learning blocks and position randomly mixed. The impact of the NHS effect on categorization should manifest in differences in responding when the critical feature appears in near vs. far positions during touch screen but not mouse interaction.

**Results.** The effect of NHS on categorization judgments was considered through analysis of RT data for all correct categorization trials beginning in the run in which the participant reached criterion, thus trimming the trials during which category learning was occurring. Mean RTs shown in Figure 2 provide some support for the presence of a NHS effect in a complex categorization task. When using the touchscreen interface, participants were 94 ms faster to categorize targets in the near vs. far condition. When using the mouse interface this pattern was reversed; participants were 41 ms slower to categorize targets in the near condition. There was a high degree of variability in the data, such that the interaction in the ANOVA only trended toward significance. However, over 70% of subjects in the touchscreen condition showed the effect in the expected direction, while the mouse condition showed participants evenly split on which condition was fastest. A follow-up study with a larger sample size and a design that allows for more trials following criterion will be carried out to confirm these trends as well as explore the effect of NHS on learning rate.

The eye tracking data remain to be analyzed. Measures of first fixation location and dwell time to each feature will be analyzed based on position and feature relevance (critical to categorization vs. irrelevant to categorization) for both touchscreen and mouse interfaces. If attentional biases are driving the trend toward a NHS advantage for categorization of targets for features in NHS during touchscreen interactions, we should see more first fixations to and longer dwell times for features in NHS for touchscreen but not mouse interfaces.

**Conclusions.** The study provides preliminary support for the impact of touchscreen use on categorization of complex images (c.f. Andringa, Boot, Roque & Ponnaluri, 2018). These findings have implications for the development of interfaces for tasks that involves image analysis (i.e. microscopy, radiology, baggage screening).

*Catherine M. Arrington (Lehigh University)*

## Computational Social Science

### G-3

#### A Closer Look at Panic Disorder Using Smartphone-Based Ecological Momentary Assessment, Wearable Psychophysiological Measurement and a Symptom-Level Approach

Background: Panic disorder is prevalent cross-nationally, with 1.7% of the population having lifetime prevalence of panic disorder, and 13.2% having lifetime prevalence of panic attacks (Jonge et al., 2016). The predominance of Panic Disorder cross-nationally suggests a need to better understand the symptomatology of patients with Panic Disorder and how symptoms change over the course of treatment. Most studies examining the efficacy of cognitive-behavioral therapy, considered the gold-standard panic treatment, for Panic Disorder have two noteworthy limitations. First, they rely on retrospective recall of symptoms over weeks or months. Further, these measures are administered in artificial settings such as a doctor's office or a research laboratory. Second, they use aggregate sum scores of Panic Disorder to assess treatment efficacy. These methods limit our understanding of how cognitive-behavioral therapy has its effect. Retrospective recall is prone to bias and sum scores mask potential differences in how individual symptoms respond to treatment. In this study, we address these limitations by using ecological momentary assessments (EMA) to examine the effects of cognitive-behavioral therapy on individual symptoms of Panic Disorder.

Methods: In this study, patients (n=20) received seven sessions of cognitive-behavioral therapy for Panic Disorder. Patients completed EMA assessments on their smartphones using the application LifeData for two weeks (14 days) at pre- and post-treatment. During these data collection periods, patients were prompted to complete the 20-item EMA questionnaire five times per day at random time intervals between 9am and 9pm. In each questionnaire, patients rated core panic symptoms, behaviors, and thoughts on a 10-point scale ranging from Not At All to Very Much. We used this questionnaire to calculate scores for five symptoms: cognitions, emotions (i.e., feeling anxious or panicky), arousal-related bodily sensations, escape behavior, and avoidance behavior. We assessed the within-subject mean and standard deviation for each symptom for each individual patient at pre- and post-treatment.

Results: We used paired-samples t-tests to compare within-subject means and within-subject standard deviations for each of the five symptom scores at pre- vs. post-treatment. All five symptom scores exhibited significant reductions in both within-person mean (mean differences 0.55-0.90, all ps < 0.04) and within-person standard deviation (mean differences 0.46-0.71, all ps < 0.01), suggesting that symptoms reduced both in severity and in variability after treatment. The largest reductions in both mean severity and variability were seen for the emotion symptom (d = 0.80 and d = 1.01, respectively). To examine more complex system dynamics and temporal instability, mean square successive difference and probability of acute change will also be presented.

Discussion: Patients exhibited a significant reduction in both the within-person mean and the within-person standard deviation of each symptom (cognitions, emotions, arousal-related bodily sensations, escape behavior, and avoidance behavior), suggesting treatment lowered both the severity and variability of Panic Disorder symptoms. Lowered within-person variance indicates that patients' symptoms become more stable after treatment. Notably, information

about the within-person variance of individual symptoms is only available with an EMA approach, illustrating its importance in understanding a patient's improvement. Similarly, we found significant differences between symptoms in their response to treatment, illustrating the value of a symptom-level approach to evaluating treatment outcomes. We can use the information gathered about how different symptoms respond to treatment to modify existing treatments to target specific symptoms. This propels us toward the possibility of developing treatments that are tailored to the individual. EMA measures should be used more widely in clinical research to allow us to develop a better understanding of precisely how treatments have their effect on Panic Disorder. Future directions for research will be discussed in the context of implications for using EMA and technology to enhance the delivery and efficiency of cognitive-behavioral therapy.

Amanda Baker (Massachusetts General Hospital)

## Empirical Studies

### H-1

#### Does the Developer Matter? Parents' Views of Mental Health Apps Developed in Academia and Industry

In recent years, the number of smartphone applications ("apps") focused on health care, including mental health care, has increased dramatically (Pohl, 2018). Even though most evidence-based treatments in mental health care originate in clinical research settings, some evidence suggests that potential users of mental health apps may have concerns about the quality of apps developed and evaluated in academia (Schueller, Neary, O'Loughlin, & Adkins, 2018). However, the extent to which these concerns translate to measurable differences in treatment acceptability has not been studied. This study evaluated the acceptability of a treatment app for youth with disruptive behavior problems; specific factors such as the work setting of the developer (e.g., academia, industry) as well as the presence of options for contact with a human therapist through the app were explored. Parents (N = 368; 48% female; M[age] = 40.25 years) with children under 18 years old completed an online survey. Approximately 23% of parents reported having a child with a formally diagnosed mental disorder. Participants were randomly assigned to one of two clinical vignettes describing a child with either (1) attention-deficit hyperactivity disorder (ADHD; n = 179) or (2) oppositional defiant disorder (ODD; n = 189). Each vignette was followed by a description of a problem-solving app designed for youth with problems like those described in the clinical vignettes. Both the clinical vignettes and the descriptions of problem-solving therapy were based on materials in published research on evaluating the acceptability of technologically novel interventions (Rabbitt, Hong, & Kazdin, 2015). Parents were randomized to one of three types of apps: (1) a problem-solving app created by academic researchers (n = 120), (2) a problem-solving app created by industry professionals (n = 125), or (3) a problem-solving app developed through a collaborative process involving both academics and industry professionals (n = 123). Each app included two versions that parents rated: (1) a version without the option of consulting with a human therapist trained in problem-solving therapy for youth and (2) a version with therapist consultation as an option. All other aspects of the app were

held constant across all conditions. Participants provided ratings of treatment acceptability (using the Treatment Evaluation Inventory; Rabbitt et al., 2015) for each version (without and with therapist consultation) for each of the three types of apps. Results from a repeated-measures ANOVA indicated that there were not differences in acceptability rating ( $F(1, 263) = 0.01, p = .97$ ) based on the clinical vignette (i.e., ODD or ADHD). Therefore, subsequent analyses did not explore vignette-based differences. A second repeated-measures ANOVA (including developer type as a between-subjects variable and degree of therapist involvement as a within-subjects variable; see Figure 1) found differences in acceptability ratings based on therapist involvement ( $F(1, 265) = 48.45, p = .000$ ), with the apps including the option of therapist consultation ( $M = 74.82$ ) rated as more acceptable than the apps without the option of therapist consultation ( $M = 71.11$ ). In spite of this difference, both apps were associated with acceptability ratings in the moderate to moderate/high range (score range: 15-105). The effect for developer type was not significant ( $F(2, 265) = 1.42, p = .24$ ), suggesting that parents viewed apps developed in different settings and by different experts as similar in acceptability. These results indicate that the possibility of consulting with a trained mental health professional is an important consideration for parents when evaluating treatment apps. Moreover, the issue of who develops the app (i.e., an academic researcher or a technology industry professional) does not influence parental views on acceptability. Given the importance of acceptability as a construct in understanding engagement in and completion of traditional treatments (e.g., Kazdin, Holland, & Crowley, 1997), this suggests that developers of apps should consider how to create opportunities for parents to consult with trained mental health professionals through apps that address disruptive behavior problems in youth. Additional work is needed to determine if a similar pattern of findings is obtained with other key mental health treatment constructs (e.g., expectations for improvement in treatment). In addition, studies are needed in which parents interact with apps directly (e.g., experiencing content, design, and features) to explore whether parental reactions differ when they are using an app rather than reading a description of one.

*Sarah Rabbitt (Oberlin College)*

## H-2

### Electronic Media and its Application in Psychotherapy: Preliminary Data

**Introduction:** A growing body of literature indicates that clinically significant behavior markers can be extracted from mobile and communication data [1,2]. However, it remains less clear whether and how such data can be incorporated into the care process, and whether they may actually have an impact in psychotherapy. Here, we present an ongoing study that assesses the impact of incorporating data from patients' electronic media (e.g. email, text messaging, social media) into therapy—referred to as Electronic Media Enhanced (EME) augmentation of standard care—when compared to Treatment as Usual (TAU) after 6 months. We describe this study design, including an adaptive approach anchored around privacy and patient control of data, and findings from our initial group of subjects. We hypothesize that both therapists and patients will report an improved working alliance and overall satisfaction with treatment after augmenting TAU with EME.

**Methods:** This study pilots a process for accessing a patient's electronic messaging content and behavior and assesses ongoing

clinician attitudes toward incorporating their patients' electronic content into their treatment approach. We plan to enroll 115 subjects at McLean Hospital (up to 100 patients and up to 15 psychotherapists) over one year; this poster will present data from all subjects who have completed the 6-month protocol at the time of presentation. All patients have a primary diagnosis of Major Depressive Disorder and/or an anxiety spectrum disorder, are actively using at least one platform for text messaging, and are meeting with their outpatient psychotherapist at least monthly.

At Visit 1, subjects are given a list of 40 words/acronyms that are commonly used to communicate negative emotions (e.g., "sad" or "FML") and are asked to identify 10 key terms that they most often utilize. During each visit, study subjects complete a 30-minute "pre-session visit" with a research assistant (RA), during which the RA reviews the subject's electronic content and messaging behavior while the subject completes mood scales (PHQ-9 and GAD-7). The electronic content/messaging review is an audit of the subject's primary messaging platform to identify instances of any of the 10 key terms and record any emotional context surrounding each instance of the term. The RA then presents this content to the clinician. Following the psychotherapy session, the clinician and subject independently complete scales assessing how they felt about the therapy session.

**Analyses:** Our primary outcome measures are clinician and patient scores on the Working Alliance Inventory and satisfaction measures, as measured by the Session Rating Scale (Modified), Clinician Satisfaction Measure, Client Satisfaction Questionnaire, and Session Evaluation Questionnaire (Patient). Secondary outcome measures will investigate whether EME is associated with improvement of clinical outcomes as measured by improvement in patient scores on the GAD-7 and PHQ-9 scales.

The results presented in this poster are not yet available; however, once data is collected, investigators will use ANOVA and chi-square tests to determine initial pre-post differences within subjects and develop regression models to identify predictors of rapport. Investigators will also utilize ANOVA to perform analysis of mediator and moderator variables that impact outcomes related to social media use.

**Results:** Results from a case study of a 77-year-old man provide preliminary evidence that this content review process is viable for reviewing electronic messaging to augment psychotherapy. Additionally, this case study gives anecdotal evidence indicating the usefulness of the electronic augmentation. Here we will present further data from all subjects who have completed the 6-month protocol at the time of the presentation.

**Practical Implications:** As technology advances, the average person is spending an increasing amount of time communicating through social media and electronic means. To ignore the magnitude of information contained in these communications would be to potentially overlook a large number of warning signs that a patient is in, or approaching, a crisis situation, which may not otherwise be detected by a clinician or willingly discussed by a patient. These data indicate not only the important clinical utility of including review of electronic media in psychotherapy but also the feasibility and willingness of both clinicians and patients towards adopting this as a regular practice.

**Conclusions:** Our findings from the subjects who have completed the protocol will indicate whether messaging behavior and content can be a useful tool in enhancing the quality of care. Developing the process surrounding the electronic content review raised complex

questions regarding privacy and safety, including concerns around third parties that communicate with the subjects. Our preliminary feasibility case study suggests that patients may be willing to share private electronic media if it can benefit their treatment.

*Hannah Heintz (Harvard Medical School/McLean Hospital)*

### H-3

#### Human-in-the-Loop AI for Analysis of Free Response Facial Expression Label Sets

**PROBLEM STATEMENT:** Facial expressions (FEs) communicate a rich variety of social, grammatical, and affective signals. However, the generally accepted set of recognizable FEs remains limited to seven basic emotion displays. One roadblock to establishing a broader and more nuanced lexicon of expressions has been the limiting nature of forced-choice study designs, the most common paradigm for investigating observer judgements of facial communication. However, there has been no consensus on an objective way to relate free responses to measures of agreement.

We present a human-in-the-loop artificial intelligence pipeline for analyzing sets of freely chosen natural language labels. The pipeline makes it possible to identify whether there is consensus on the signal value of an FE, and if so, which label best classifies it. Unlike the manually applied relatedness heuristics of prior free-response analyses, it scales to process very large datasets. We validate our approach in two stages: comparison of label relatedness scores output by five computer algorithms to human raters' label synonymy scores; measurement of pipeline results against those obtained through manual processing of data from emotion and facial expression recognition studies.

**PIPELINE DESCRIPTION:** Our pipeline employs a 3-step process for determining whether an FE is recognizable, and if so assigns a representative label. **Assign relatedness scores:** All pairs of labels from a set are assigned a real number score from -1 to 1 (with 1 being an exact match) by calculating the cosine similarity of their word vectors. **Clustering:** Hierarchical agglomerative clustering determines the number of semantic clusters present in a set. If a single cluster does not contain at least 75% of the labels, the set is deemed incoherent, and the FE is considered not reliably recognizable. **Label weight calculations:** Relatedness scores are summed, and the label with the greatest weight is considered the representative label.

**COMPARISON STUDIES:** Synonymy Scoring Comparison. Fifteen judges assigned synonymy scores ranging from 0 (exact opposites) to 10 (same or interchangeable) to 156 word pairs. We tested five algorithms that employ different bases for scoring relatedness on these pairs, and calculated agreement between algorithms and human judges using Krippendorff's  $\alpha$  (Table 1). Numberbatch performed best ( $\alpha = 0.561$ ). While it did not meet the 0.6 threshold considered good agreement, empirical performance is on par with the results of manual processing.

**Comparison with Manual Analysis.** Early free response studies did not involve rigorous, repeatable designs. Most related subjects' responses to predetermined lists of predicted labels and provided transformed data rather than full response sets. We processed data from three studies using our pipeline, and results were compared with the source.

**Haidt & Keltner:** Undergraduates (N=40) freely labeled 14 emotion FEs. A complete response list is available only for the expression

of contempt. Researchers clustered responses with target emotions based on synonymy and used binomial tests of significance to determine recognizability. They found that participants did not recognize the expression of contempt.

In our pipeline this label set passed cluster analysis and was deemed reliably recognizable as "annoyed" (Figure 1, Table 2). Rosenberg and Ekman similarly found that "annoyed" was provided as the modal label in a free response recognition task.

**2.2.2 Oster & Kyi:** Oster and Kyi's unpublished study tested whether untrained observers could accurately identify the emotions shown in spontaneous FEs produced in real-life emotion-arousing situations. Twenty-five undergraduates used their own words to identify the emotions shown in 38 slides of facial expressions taken from TV news and documentaries, including FACS-coded expressions of seven "basic" emotions and hypothesized expressions of distress and resignation.

Observers' free responses were categorized into the nine categories by a thesaurus search for synonyms and a point system for rating the synonyms' closeness to predicted emotions, including full points for relevant situations or behaviors (e.g., "crying"). Summed points were used in binomial significance tests.

Our AI pipeline agreed with results for 34 of 38 slides. Two images portraying distress and one of disgust that failed binomial tests passed our clustering threshold. One hypothesized distress face was categorized as sadness, the other as fear. The AI pipeline failed an anger expression that passed the binomial test, demonstrating how our pipeline penalizes the magnitude of disagreements, unlike other methods (Figure 2).

**Shaver:** Shaver had 100 undergraduate students sort 135 emotion words into freely chosen groups based on similarity, then developed a dendrogram visualizing six primary semantic clusters (love, joy, surprise, anger, sadness and fear) and 25 subclusters from the groupings.

The dendrogram generated by our pipeline reproduced Shaver's six primary clusters. Of 25 subcluster labels, 18 fall within the expected primary clusters. Some additional clusters in our dendrogram indicate emotion categories that may be of interest, such as loneliness

*Crystal Butler (New York University)*

### H-4

#### Teachers Leveraging Game-Based Implicit STEM Learning in Classrooms

**Introduction:** Game-based learning introduces novel forms of formative assessment (Gee, 2013; Shute et al., 2013). Researchers have made use of data logs generated from digital games to study the development and measurement of students' tacit and implicit knowledge (Authors, 2015a; 2017). Implicit knowledge has been argued to be foundational in explicit learning (Polyani, 1966). Implicit learning in games, however, does not ensure explicit learning in the classroom and the connection between the two must be facilitated to support successful transfer of knowledge (Hattie & Yates, 2013). In our previous research with physics learning games, we found that teachers' use of bridging activities to connect game-based learning to classroom content was a significant factor in predicting post-assessment scores while controlling for pre-assessment scores (Authors, 2014; 2017). Our current research examines this bridging model with the

construct of computational thinking (CT) (Wing, 2006). In a national classroom implementation study of the logic puzzle game Zoombinis, 1,273 students in elementary and middle school classes and teachers were provided with classroom materials designed to help bridge game-based knowledge with formal instruction. This paper focuses on the analysis of written 1,051 teacher logs to determine the extent and nature of bridging in those 57 classes.

Bridging in Zoombinis. Zoombinis (Authors, 2015b) is an award-winning, educational game targeting ages 8 through adult, consisting of a series of challenging logic puzzles designed to develop mathematics concepts essential for computer programming and data analysis, such as sets, logical relationships, dimensions, mappings, sorting, comparing, and algorithms (Figures 1a-1c).

We designed online and offline classroom activities, video clips, and discussion prompts about computational thinking within gameplay to support teacher bridging. Teachers decided how to implement Zoombinis into their instruction. This study explores two research questions:

1) In what ways do Zoombinis classrooms vary in their use of bridge materials, game-centered pedagogy, computational thinking instruction, Zoombinis gameplay, and teacher, classroom, and school demographics? 2) Do these classroom-level characteristics cluster in any meaningful way? (e.g., proportion of gameplay relative to bridging)

Methods: Data Collection and Teacher Features. Thirty-six US teachers with 57 3rd-8th grade classes participated in this study during the 2017-2018 academic year. Using NoviSurvey software, teachers kept daily logs of computational thinking activities for each study class and reflected on their practice before and after the study. Each teacher was supported by an EdGE researcher during their implementation of Zoombinis.

In addition to teacher logs, we collected the following types of data from teachers: surveys about teachers' prior game usage, teacher, classroom, and school demographics, and 14 classroom observations.

We are currently creating four major categories of features based on these data sources. Sample features include: 1. Computational Thinking Instruction: Total time spent on any computational thinking activities, Percentage of time on programming activities; 2. Game-centered pedagogy: Total time spent on Zoombinis activities, Percentage of time spent on Zoombinis individual gameplay; 3. Zoombinis gameplay: Total amount of gameplay for all students in class, Total number of puzzles tried by at least 50% of students in class; 4. Teacher, Classroom, and School Demographics: Prior experience using games in the classroom, >50% students in free/reduced lunch within the school.

Analyses. We are analyzing 1,051 daily, pre-implementation, and post-implementation teacher logs as well as 36 teacher application surveys to create these features. This paper will present a descriptive analysis of how Zoombinis instruction varied across 57 classrooms. These results will be combined with 14 classroom observations that focused on how teachers taught computational thinking skills and how they bridged these concepts to other subjects.

With these features, we will apply k-means, a clustering algorithm, which partitions data into k distinct clusters by assigning each observation to the nearest cluster centroid. Potentially meaningful clusters might mirror what was found for Game Only vs. Bridging classrooms in our prior work with a high school physics game (Authors, 2014). Here, Game only classrooms would be when

teachers reported gameplay but little use of the bridge materials. Bridging classrooms, in contrast, would be those where teachers used Zoombinis materials and other activities to connect gameplay with CT classroom instruction.

Discussion and Conclusion. Findings of this study will provide insight on what bridging looks like in the context of a computational thinking game, including the distinct ways teachers implemented bridging activities in the classroom. Consistent with our previous work on bridging in two physics games (Authors, 2014; 2017), features from this study will be used for multilevel analyses to examine the relationship of Zoombinis bridging activities on changes in students' pre- and post-assessments scores on computational thinking. Our study represents a first step towards replicating our bridging model to a broader context of games and computational thinking instruction. Overall, our research has potential implications for incorporating games in teacher professional development and supporting teacher bridging to improve explicit STEM learning.

Ma. Victoria Almeda (TERC)

## Graduate Education / Professional Training

### I-1

#### Toward a User Experience Research Training Program within a Psychology Program

"User experience" (UX) is defined as a person's total experience with technology (Norman & Nielsen, n.d.). UX researchers come from a number of backgrounds including but not limited to, computer science, psychology, visual arts and technical writing. Key characteristics of UX researchers include an interest in the well-being of users, a willingness to work in interdisciplinary team-based environments, tolerance for changing technology, comfort working with ill-defined problems, and a love of problem solving (cf. Leventhal and Barnes (2008). The basic skills needed for a successful career in UX research include skills that psychology students gain through their psychology courses: Ability to work on a team; Understanding of research methods; Understanding of data collection and analysis; Ability to communicate verbally and visually

UX researcher jobs are in high demand and offer exciting careers for psychologists (Baker-Burgoon, 2016; Glassdoor, 2018; Kremer, 2013; Money-CNN, 2017; Nielsen, 2017). But training in a psychology program does not impart all of the skills that are needed to become a UX researcher. A successful UX researcher needs three additional skills: Ability to apply theories and practice of psychology to UX research; Understanding of the processes of UX design and engineering; Understanding of the specialized methods of UX research

Historically, training programs for careers in UX research have been housed in computer science and/or information science programs (Hewett et al., 1992); as such, psychology students may not be aware of or able to access these programs. We believe that psychology students should have opportunities to train within their home discipline of psychology. In this poster, we describe three courses that are designed to prepare psychology students for careers in UX research; these courses (described below), depending on their depth, could be appropriate for either graduate or undergraduate students

Applied cognition in UX. The purpose of this course is to introduce psychology students to applied areas of psychology, with a

focus on how the study of psychology and its sub-disciplines (cognition, clinical, social, etc.) are applicable to UX research. This course is intended to develop the skill: Ability to apply theories and practice of psychology to UX research. Learning objectives: Students should: 1. Understand the relationship between UX research and approaches/facets/theories of psychology. For instance, students would learn how some aspects of UX research might be informed by an understanding of clinical practice; 2. Be able to apply psychological theory to specific problems in UX. For example, students would learn how to apply their knowledge of perceptual organization constructs to the design and assessment of the visual organization of a web page.

Usability Engineering. This course introduces psychology students to the processes and methodologies of usability engineering. This course is intended to develop the skill: Understanding of the processes of UX design and engineering. Learning objectives: Students should: 1. Become familiar with popular approaches to and the language of usability engineering; 2. Become familiar with the key stages of design/engineering processes and understand how a UX researcher contributes to each stage of the processes. For example, early in the process of developing a technology solution, a UX researcher might identify and analyze user characteristics of the target user. Then during design, the UX researcher would assess how well the design matched those user characteristics; 3. Have appreciation for and experience with solving problems on an interdisciplinary teams (e.g. computer science students, visual designers, etc.)

Applied research methods in UX research. This course serves as an additional research methods course, to follow a basic research methods course in a psychology curriculum. Unlike a basic research methods course, the focus of this course is on methods specifically involved in UX research. This course is intended to develop the skill: Understanding of the specialized methods of UX research. Learning objectives: Students should: 1. Gain an appreciation and understanding of the research process and methods specific to solving problems in UX research; 2. Understand how to choose research methodologies while conducting UX research in a given situation. For example, a UX researcher might be called upon to assess a particular technology solution with a very short time to market window. A traditional laboratory experiment might be infeasible. Students would learn how to choose the most effective research methods, given the constraints of the development environment; 3. Feel prepared to use their knowledge in applied settings. Following these three courses, students might pursue a fourth practicum course and/or industrial internship.

In summary, UX research is a growth field with great career potential for psychologists. In reviewing the courses above, we hope to begin a discussion of UX research training within psychology programs.

*Samuel Jaffee (Google)*

## Health / Well-Being

### J-1

#### **An Experimental Prosocial Intervention to Reduce Cyberbullying in YouTube Comments: Mixed Methods Findings**

Cyberbullying, defined as “willful and repeated harm inflicted through the use of computers, cell phone, or other electronic devices” (Hinduja & Patchin, 2009), is a worldwide emerging public health

concern (Tokunga, 2010). According to research findings, 41% of the people in the U.S. reported experiences of some form of online harassment, and 66% reported online harassment as a serious problem (PEW Research Center, 2017). More specifically, cyberbullying is consistently associated with various deleterious psychological and behavioral outcomes including anxiety, depression, substance use, suicidal ideation (Kowalski, Giumetti, Schroeder, & Lattanner, 2014). In turn, more prevention science scholarship is being focused on prevention of cyberbullying and its adverse consequences (Van Geel, Vedder, & Tanilon, 2014).

Previous literature findings about traditional bullying have demonstrated that the action or inaction of bystanders can alter the consequences of bullying (Polanin, Espelage, & Pigot, 2012). Bystander intervention happens when an individual that witnesses bullying decides to intervene (Polanin, Espelage, & Pigot, 2012) and it has been found to be an effective way of mitigating other forms of bullying (Salmivalli, 2014). Yet, literature findings indicate that the likelihood of bystander intervention is very low in cyberbullying incidents (Huang & Chou, 2010). Thus, there is an emerging focus on promoting bystander intervention in bullying in digital platforms as a way of preventing cyberbullying (DiFranzo, Taylor, Kazerooni, Wherry, & Bazarova, 2018). Although, research findings indicate that there are some strategies that are commonly observed in individuals’ way of intervening when they are bystanders in bullying incidents (Freis & Gurung, 2013; Shultz, Heilman, & Hart, 2014) little is known about the effectiveness of the intervention strategies that are used by bystanders in the virtual world.

The present study is part of a project funded by the National Science Foundation, that aims to contribute to building a healthier online ecosystem by creating an internet chat bot (BystanderBots) that automates the detection of toxic words and intervenes in cyberbullying incidents by automatically posting prosocial comments as if it were a real person. Our study was divided in 5 phases: (1) YouTube data collection and cyberbullying annotations; (2) experimental development of a machine learning detector of abusive language; (3) data collection and annotation of prosocial comments found in the wild; (4) generation of new prosocial comments; and (5) testing prosocial comments generated using a randomized experimental design to evaluate the efficacy of prosocial comments to reduce the incidence of cyberbullying.

Previous Results: During Phase 1, human annotators assigned cyberbullying labels (1, Yes; 0, No) to 11,540 YouTube comments extracted from 253 different channels publishing anti-feminist content. Annotators classified 27.5% of the comments as abusive and 20.2% of the sentences were identified as containing cyberbullying language. During Phase 2, the team developed novel neural network techniques to detect abusive language. The detector uses a novel embedding-based neural network approach to disambiguate the different senses in which toxic words can be used, since many swear words can be used both in an aggressive and a friendly sense. The embedding-based iterative approach achieved comparable performance to human annotators and allowed us to compile a quality profanity lexicon without reliance on external linguistic resources. During Phase 3, the team collected a sample of prosocial comments found in the wild which is used to inform the following phases.

The aim of Phase 4, the current stage, is to create a set of prosocial comments that will be used by BystanderBots in the automated intervention response and in Phase 5 the effectiveness of the created prosocial comments on mitigating cyberbullying will be tested.

To do that, the present study aims to generate prosocial comments that are likely to have an impact on mitigating cyberbullying as well as increasing other individuals' willingness to intervene on YouTube comment threads. In the light of previous literature findings, six categories of prosocial comments (empathy-provoking, promoting constructive conversation, changing the subject, addressing the topic not the person, removing yourself from the audience, and complimenting/telling jokes) were created. The corpus of prosocial comments will be posted to 100 comment threads on YouTube. Another set of 100 comments will be used as the control condition. The prosocial intervention group will be compared with the control group using the number of cyberbullying comments posted within a period of 30 days from the publishing date of the original prosocial comment and a matching comment from the control group. Additionally, the follow-up comments posted after the prosocial comments will be coded qualitatively in different domains (e.g. effectiveness, appropriateness, unintended effects). The mixed methods results will allow us to understand the complexity of online toxic discourse and to identify and describe which categories of comments have the potential to reduce the incidence of online cyberbullying. The results from Phase 4 and 5 will be available during the next few months, and the presentation will be based on those findings.

*Cagil Torgal-Askoy (University of Florida)*

## J-2

### **Building Post-Disaster Resilience in Puerto Rico: Developing Offline Digital Tools for Mental Health Support Service**

According to public health officials, there are warning signs of a full-fledged mental health crisis in Puerto Rico after Hurricane Maria, and much of the population is showing symptoms of post-traumatic stress. The rising rates of depression and suicide are among the most critical problems facing communities in Puerto Rico since Hurricane Maria.

While mental healthcare resources are available, mental health screening and diagnosis can create a barrier to patients receiving the treatment they need. In the aftermath of Hurricane Maria, primary care physicians (PCPs) are often the first point of contact for individuals who are seeking medical assistance for physical and mental health concerns. While waiting to see the PCP, patients may complete standard diagnostic questionnaires (PHQ2, PHQ9, and GAD7) for mental health screening. The results of these questionnaires indicate whether or not the patient should receive a referral for mental health treatment. These primary care physicians and their staff are often overworked, and may not have the time to manually calculate the score of the questionnaires, or expertise to face the challenge of identifying the most relevant referral to mental health care services.

To address this problem, our team is working with collaborators at hospitals and Universities in Puerto Rico to develop an offline digital application to administer the diagnostic questionnaires (PHQ2, PHQ9, and GAD7), calculate the patient's score, and offer referrals. If the score of the questionnaires indicates the patient should receive mental health treatment, the application will output a list of multiple mental health service providers that fit the patient's needs based on information they have entered (such as insurance, home address, and mode of transportation). The results for referral are populated from a backend database of mental health care providers in Puerto Rico. All information is erased from the tablet or device before the device is passed on to another patient.

This application will offload work from the PCPs and staff, as they no longer need to invest time manually calculating diagnostic scores or searching for an appropriate mental health service referral. The offline database of mental health service providers is a resource not currently available at most PCP locations. In the future, we aim to expand this project to include specialities of the mental health service providers as part of the referral list (e.g., veterans, adolescents, LGBTQ, etc) and distribute the application to more primary care physicians, allowing even more patients to receive the mental health care services they need.

*Annie Bruns (ATLAS Institute, University of Colorado, Boulder)*

## J-3

### **Fear of Missing Out Mediates Relations Between Problematic Smartphone Use Severity with Depression and Anxiety Symptoms in Chinese Young Adults**

Recently, studies have examined mediating psychological constructs accounting for relations between both depression and anxiety with problematic smartphone use (PSU) severity. An important psychological variable that is implicated in these relationships is the fear of missing out (FOMO) on rewarding and pleasurable experiences in one's social network. FOMO has been linked with levels of PSU, but only in Western samples; thus we know little about this research in Asian participants. And in fact, FOMO's relationship with PSU may be different in Asian participants such as Chinese residents because of different social rules in China that emphasize social resource exchange. The aim of the current study was to analyze FOMO as a possible mediator in these relationships. We used the Interaction of Person-Affect-Cognition-Execution (I-PACE) theoretical model to conceptualize our study. I-PACE proposes personal predisposition variables, and affective and coping mediators that account or problematic internet use. We hypothesized that 1) FOMO should be related to PSU severity; 2) FOMO should relate to smartphone use frequency; 3) FOMO should mediate relations between depression and PSU severity; and 4) FOMO should mediate relations between anxiety and PSU severity. We recruited 1,034 Chinese undergraduate students in 2018 via a Chinese web-based survey that measured smartphone use frequency, PSU (Smartphone Addiction Scale-Short Version), depression (DASS-21 Depression Scale), anxiety (DASS-21 Anxiety Scale) and FOMO (FOMO Scale). We used structural equation modeling (SEM) to test our hypotheses, using weighted least squares estimation with a mean- and variance-adjusted chi-square. We tested mediation using the Delta method, taking the cross-product of direct path coefficients with 1000 bootstrapped replications. SEM results demonstrated that FOMO was significantly related to smartphone use frequency and PSU severity. FOMO significantly mediated relations between anxiety and both smartphone use frequency and PSU severity. FOMO did not account for relations between depression and smartphone use/PSU. This is one of the first studies testing FOMO in relation to PSU severity among Asian participants. FOMO may be an important variable accounting for why some types of psychopathology (e.g., anxiety) are associated with PSU.

*Jon Elhai (University of Toledo)*

## J-4

**Improving the Prediction of Subjective Stress from Physiological Markers in Virtual Reality**

Stress can be understood in multiple ways (Jamieson, Nock, & Mendes, 2012). The present work focuses on the relationship between the physiological component of stress (nervous system activity) and conditions under which it predicts its subjective component. Multiple theories elucidate the relationship between the physiological component of stress and the subjective component of stress. The James-Lang Theory and the two Factor Model of Emotion both predict that the subjective component of stress is a function of the physiological component of stress. Furthermore, both theories predict that trait interoceptive sensitivity, the degree to which an individual characteristically optimizes their brain's reception of afferent bodily signals, should interact with the relationship between the subjective component of stress and the physiological component of stress (Dunn et al., 2010a). Similarly, the Theory of Constructed Emotion predicts that the subjective component of stress is a function of the physiological component of stress and that trait interoceptive sensitivity would interact with the concordance between the subjective and physiological components of stress (Barret, 2017).

The current study aims to situate these theoretical bodies of knowledge in the understudied context of virtual reality. Specifically, we use virtual reality to present situations which greatly resemble real-life phenomena, and examine the physiological component and the subjective component of stress when individuals face highly evocative stimuli, simultaneously assessing the implications of individual differences in interoceptive sensitivity on the concordance between the subjective component of stress and the physiological component of stress. Furthermore, in an exploratory phase we investigate a proof of concept artificial intelligence algorithm to predict subjective stress experiences from the physiological markers of stress.

To explore these ideas in the context of virtual reality experiences, the present work focuses on highly evocative video games which allow participants to interact with audio stimuli. The video game in the present study, *Resident Evil 7 Biohazard*, contains a forested environment, which represents a calming environment, as well as a haunted house, which represents a stressful environment (PlayStation, 2018).

Prior to immersion in virtual reality, one hundred forty-four participants completed the Schandry heartbeat perception task, a measure of trait interoceptive sensitivity (Dunn et al., 2010b). Next, participants navigated the virtual reality environments via a VR headset and a game controller, while their heart activity was simultaneously recorded by an electrocardiogram. Periodically, an experimenter asked participants questions about their alertness level, stress level, and fear level, to measure the subjective component of their stress while participants explored a calming environment and then a stressful environment or vice-versa.

We calculated participants' heart rate (HR) and heart rate variability (HRV) as parameters of their physiological component of stress for estimation of subjective stress. We hypothesized that artificial intelligence and virtual reality will facilitate stress research by aligning the physiological component and the subjective component of stress and by allowing physiological markers of stress reactivity to be used to measure subjective stress by proxy, respectively. In a separate part of the study we measured interoceptive sensitivity via the Schandry heartbeat perception task, aiming to explore whether

overall alignment of physiological and subjective stress is greater among people who exhibit high levels of trait interoceptive sensitivity (Dunn et al., 2010a). In the exploratory part of the project we examined the utility of the physiological markers of cardiovascular activity during the game for predicting the subjective stress states and the direction of stress trajectory in neutral and emotionally-evocative episodes of the game. To this end, we used a naïve Bayesian classifier and a supporting vector machine algorithm (Glick, Jenkins, Nettles, Hitchings, & Davies, 2006) to train a model allowing for prediction of subjective stress based on physiological markers. Next, in the confirmatory part of the project, we compared the stress ratings that the artificial algorithm predicted for individuals who achieved high scores and for individuals who achieved low scores on the trait interoceptive sensitivity measure to those individuals' respective heart rate measures in order to assess whether the AI model could reproduce the finding that the overall alignment of physiological and subjective stress is greater among people who exhibit high levels of trait interoceptive sensitivity, as well as testing boundary conditions of this finding (e.g., moderation by different levels of stress).

*Ethan Fischer (University of Waterloo)*

## J-5

**Keeping the Spark Alive: How Using Technology for Cybersex is Associated with Couple Satisfaction in Long-Distance Relationships.**

Background: Technology is important to many romantic relationships, as it provides an avenue of communication which is key to developing partner attachment (Schade, Sandberg, Bean, Busby, & Coyne, 2013). The use of technology may be especially relevant to long-distance relationships (LDR), as lack of proximity can be a barrier to maintaining fulfilling, intimate relationships (Stafford, 2011). With a growing landscape of communication technologies available to individuals in LDR, it is important to expand knowledge of how long-distance couples stay connected physically and emotionally, and how their use of different mediums of communication is associated with relationship satisfaction (Fincham, Rogge, & Beach, 2006). Specifically, it may be important to understand how individuals in LDR use technology to engage in sexual communication—or more colloquially, cybersex. Engaging in cybersex is associated with increased relationship satisfaction (Neustaedter & Greenberg, 2012). Simultaneously, past work on LDR individuals reports unique associations among relationship and health variables by biological sex, e.g. LDR men reported stronger associations between relationship satisfaction and several indices of health, than do LDR women (Du Bois, Yoder, Ramos, Grotkowski & Sher, 2019). We examine cybersex behaviors and their association with relationship satisfaction, by sex, in a sample of LDR dyads. We predict cybersex behavior engagement will not differ by sex; will relate positively to relationship satisfaction across sexes; and that these associations will be stronger in LDR men than women (Du Bois et al., 2019). Methods: Participants (N = 180, or 90 LDR dyads; 18 years or older; English speaking; in LDR for 3+ months) completed a paid online survey on relational experiences and health. Cybersex behaviors were assessed through an internally created scale that assessed frequency of technology use in cybersex, e.g. phone calls, texting, and video chat. Relationship satisfaction was assessed via the Couples Satisfaction Index (Funk & Rogge, 2007). We calculated a total score representing relationship satisfaction,

and used a total score for frequency measures of technology used in cybersex behavior. We first report descriptives of the sample and key study variables. Next, we report t-test results comparing men's and women's mean scores for total cybersex behavior. Finally, we report regression associations between key study variables, among each sex. Given the dyadic nature of these data, we used this analytic approach to avoid data dependency confounds. Results: Overall, participants reported highest levels of using technology for cybersex through the use of texting ( $M=1.81$ ,  $SD=1.38$ ), followed by talking over the phone ( $M=1.63$ ,  $SD=1.42$ ), and video chatting ( $M=1.52$ ,  $SD=1.49$ ), sending videos over the computer ( $M=1.43$ ,  $SD=1.26$ ) and finally social media ( $M=1.12$ ,  $SD=1.36$ ). Total cybersex behavior did not differ by biological sex, ( $F(1, 178) = .128$ ,  $t = .069$ ,  $p = .290$ ). In males, cybersex behavior positively predicted relationship satisfaction, ( $F(1,88)=11.25$ ,  $p=.001$ ) with an  $R^2=.113$ ; in females only, no significant association was found ( $F(1,88)=4.666$ ,  $p=.031$ ) with a  $R^2=.050$ . Conclusion: Men and women in LDR seem to engage in cybersex with similar frequency. Engaging in cybersex behaviors is associated with higher levels of relationship satisfaction among men, but not women, in LDR. This is a replication of past findings that indicate LDR men report more significant associations between relational variables than do LDR women. These findings necessitate additional research to clarify associations between cybersex behaviors, relationship variables, and individual health, among those in LDR—a population that is growing both in prevalence and cultural significance. These findings also have applied implications: Those who produce technology can use these findings to collaborate with individuals in LDR, to create technology that facilitates relationship-building in LDR. Practitioners who work with individuals in LDR, or LDR couples, can use these results to inform assessment and treatment of LDR individuals seeking healthcare.

*Humza Khan (Illinois Institute of Technology)*

## J-6

### Measuring Smartphone Dependence Among College Students: Development and Validation of a Psychometric Scale

In the U.S., as high as 86% of age 18-29 individuals own a smartphone (1) and 89% of age 13-17 have or have access to a smartphone (2). Smartphone can be used to enhance productivity, for example, students may use their phones to manage their study timetables, access textbooks and learning materials, and communicate with their teachers. On the other hand, smartphone use might related to some potential problems, for example, dangerous use while driving or crossing the street (3), prohibited use during classes (4), sleep interference (5), and addiction-like symptoms (6). There has been a heated discussion on whether "smartphone addiction" should be taken as a clinically relevant disorder. However, the conceptualizations of the existing scales for smartphone addiction or dependence are vary and unclear (7) and none was developed in reference to the nine criteria of Internet Gaming Disorder (a similar technology-device related phenomenon) listed in the fifth edition of the Diagnostic and Statistical Manual for Mental Disorders (DSM-5). This study aimed to develop a scale for measuring smartphone dependence and evaluate its reliability and validity.

Participants were 1009 undergraduate students from a Midwestern university. In Fall 2018, they responded to the 18-item Smartphone Dependence Scale. The items were developed in

reference to the nine criteria of Internet Gaming Disorder listed in DSM-5. Two items were developed per each criterion. For example, an item "I use my phone to escape from negative feelings (such as stress, loneliness, and boredom)" was developed for the criterion "escape or relief from a negative mood"; an item "When not using my smartphone for some time, I find myself constantly thinking about checking my phone" was developed based for the criterion "preoccupation"; an item "I have attempted to spend less time on my smartphone but am unable to do so" was developed for the criterion "unsuccessful attempts to stop or reduce"; and an item "I secretly check my smartphone when others expect me to be doing something else" was developed for the criterion "deception". Responses were elicited based on a 5-point scale (1=strongly disagree to 5=strongly agree). In addition, the study participants provided self-reported data on their time spent on smartphone, sleeping, and physical activities. Their level of anxiety was assessed by the 7-item Generalized Anxiety Disorder scale (GAD-7) and their level of depressive symptoms was assessed by the 20-item Center for Epidemiologic Studies Depression (CES-D) scale.

Factor analysis was employed to test the factor structure and measurement invariance of the scale. Exploratory factor analysis found seven latent factors: (1) preoccupation, (2) withdrawal symptoms, (3) loss of interest in other hobbies and activities, (4) deception, (5) escape or relief from a negative mood, (6) tolerance and unsuccessful attempts to stop or reduce, and (7) continue despite problems and risk/lose relationships. Each of these factors mirrors its corresponding DSM-5 criteria, except the items for criteria "tolerance" and "unsuccessful attempts to stop or reduce" were clustered into one factor, so as the items for the criteria "continue despite problems" and "risk/lose relationships". Confirmation factor analysis found that this seven-factor model fits the data well,  $\chi^2=415.42$ ,  $df=114$ ,  $RMSEA=0.05$ ,  $CFI=0.94$ . Reliabilities of the seven subscales were satisfactory (Cronbach alphas:0.67-0.76). Measurement invariance of the scale across gender and majors was supported. Correlational analysis found that higher levels of smartphone dependence indicated by the sum of the 18 items) were associated with more time spent on smartphone, less sleeping, less physical activities, and higher level of anxiety and depressive symptoms ( $ps < .05$ ).

To conclude, empirical evidence was found for supporting Smartphone Dependence Scale as a reliable and valid instrument for measuring smartphone dependence. The instrument opens up the feasibility of researching the effect of smartphone dependence on college students' functioning and development, which in turn, provides research evidence for facilitating the discussion of whether smartphone addiction should be taken as a disorder.

*Angela Chow (Indiana University Bloomington)*

## J-7

### The Effects of Ratio and Escalating Reinforcement Schedule on Walking Behavior among College Students

Physical activity is an essential condition for healthy life, and it is effective in physical health (Hills, Andersen, & Byrne, 2011) as well as mental health (Rethorst & Trivedi, 2013). Therefore, the need for intervention to increase physical activity has been highlighted. The theory of operant conditioning in behaviorism describes behavior occurrence as a three factors contingency - antecedent, behavior, and consequences (Ferster & Skinner, 1957). On the basis of this theory,

it is explained that the occurrence of a behavior in the future can be increased by “reinforcement”. Including “reinforcement” system as an intervention strategy to facilitate physical activity has been proven to be effective across various groups, from children (Finkelstein, Tan, Malhotra, Lee, Goh, & Saw, 2013) to the elderly (Farooqui, Tan, Bilger, Finkelstein, 2014). Reinforcement schedule, a set of rules that determine the conditions under which behaviors are reinforced, has gained attention as an effective strategy in the literature. There are four types of basic reinforcement schedules based on ratio or interval criteria and fixed or variable criteria (Ferster & Skinner, 1957), and previous studies comparing the effectiveness of these schedules in behavior change consistently reported that variable-ratio is the most effective schedules (Chaudhri et al., 2007). Recently, escalating schedule which is characterized by providing an incremental reward for the continuous behavior achievement, while initializing the reward to the lowest level in case of failure (Roll & Shoptaw, 2006), has been studied as an alternative. This modified form of reinforcement schedule has been found to be an effective strategy for changing many health-related behaviors (Cassidy et al., 2018). However, there are only limited studies that have so far applied a reinforcement schedule to physical activities and compared its effectiveness. This is due to the practical difficulties with continuous behavior monitoring and immediate reward delivery contingent to the achievement of behavior, which is essential for effective application of reinforcement schedule. Fortunately, mobile technology developed with 4th industrial revolution solved these problems through in-device sensing technology (Dallery, Kurti, & Erb, 2015; Fanning, Mullen, & McAuley, 2012). Therefore, the purpose of this study is to verify the most effective reinforcement schedule in increasing walking behavior by using mobile devices.

A 7 weeks long experiment of two (fixed vs. variable ratio) by two (escalating vs. constant) by two (pre vs. post) design was conducted on 88 male and female college students. First, all participants visited the laboratory and responded to the Barratt Impulsiveness Scale-11 (BIS; Barratt, 1959), Self-report Habit Index (SRHI; Verplanken & Orbell, 2003), and International Physical Activity Questionnaire-Short Form (IPAQ-SF; IPAQ Research Committee, 2005). The baseline was measured by counting the number of steps for the next five days. Then all participants were randomly assigned to the four conditions; Escalating-Fixed Ratio (E-FR; N = 23), Escalating-Variable Ratio (E-VR; N = 20), Constant-Fixed Ratio (C-FR; N = 22), and Constant-Variable Ratio group (C-VR; N = 23). All participants were provided with cash-changeable points for 6 weeks through mobile devices based on their steps, but the points providing criteria and quantities of points were varied depending on the condition. In the FR, points were given every time the participant achieved a fixed number of steps. In the VR, on the other hand, points were given every time the participants achieved the certain number of steps—the number of steps varied first goal after the other. In the Escalating, if the time interval between achieving walking goals was less than 3 days, the points provided were increased by 100. Otherwise, 100 points were given. Whereas, participants in the Constant were always given 1,050 points each time they met their walking goals. 6 weeks later, all participants responded to the same questionnaires that were used in the first day.

Three-way mixed measure ANOVA was conducted to analyze the effect of intervention, FR/VR and Constant/Escalating in increasement of steps. Results showed the significant main effect of the intervention based on reinforcement schedules, but the all 3way

and 2way interactions were not significant. On average, the every-day step counts of participants significantly increased about 1,500 after intervention regardless of the type of reinforcement schedule. And Two-way ANOVA was conducted to analyze the differences in the amount of points provided between groups. Results showed that mean amount of points provided in Escalating groups was approximately 50% less than that of Constant groups, and the difference was statistically significant. These results indicate that the escalating reinforcement schedule is more efficient than other schedules since it promoted walking behavior with less rewards. The implication of this study is that reinforcement schedules can effectively promote many areas of mobile healthcare. Especially, escalating reinforcement schedule demonstrates efficient and effective way of promoting physical activity. Further studies have to make an effort to incorporate theoretical bases when developing technology-based intervention of physical activity and other mobile health care.

*Changseok Lee (Yonsei University)*

## J-8

### **Wellbotics: Empathic Design of Robotics-Based Psychoeducational Intervention for Children, Teens, and Families Facing Cancer**

Client engagement is essential. Engagement is especially challenging when establishing relationships between therapists and children, teens, and families. According to Landreth and Bratton (1999), “The play therapy process can be viewed as a relationship between the therapist and the child in which the child utilizes play to explore his or her personal world and also to make contact with the therapist in a way that is safe for the child.” Play therapy strategies have been shown to increase client engagement (Levy, 2008; Tompkins, 2018). Play therapy featuring video games has shown promise in leading to client engagement among those that are comfortable with electronic gaming (Horne-Moyer, Moyer, Messer, & Messer, 2014).

Advances in robotics, the branch of technology that deals with the design, construction, operation, and application of robots, has led to a line of related digital toys and games. Robotics construction kits allow users to design, construct, code, and operate autonomous robots. Using these kits in recreational, educational, and competitive contexts has been shown to engage participants while building community and increasing problem solving skills (Huskens, Palmen, Van, Lourens, & Barakova, 2015; Huijnen, Lexis, Jansens, & de Witte, 2016). Users interact with these kits under varying nomenclature including making, coding, and robotics. When these kits are used to build, code, and operate robots, users encounter many of the therapeutic aspects of art, block based play, narration, and electronic games shown to be successful in engaging clients. Accordingly, the author proposes that robotics, making, and coding can be used as a programmatic intervention as it applies to children who have family members who have been diagnosed as terminally ill or children facing other types of family trauma. This is the focus of the presentation.

**Purpose.** This poster presents findings from the Planning and Prototyping cycles of the Empathic Design process of Wellbotics. Wellbotics is a psychoeducational intervention which applies play therapy in the form of making, coding, and robotics within cancer care communities. Wellbotics provides children, teens, and families with information about experiences related to a family cancer diagnosis. Wellbotics also provides clients with language that enables them to access and communicate about information and support options, in a noninvasive, non threatening manner.

The design methodology utilizes a five cycle iterative empathic design process with three participant groups in each cycle. The cycles are: Planning, Prototyping, Testing, Analyzing, and Debugging. Each cycle includes five phases also named: Prototyping, Testing, Analyzing, and Debugging. The analyzing and debugging phase of each cycle informs the planning and prototyping phase of the next cycle, hence the iterative nature of the design process.

This research is not only iterative, but also collaborative. Each iteration summons a qualitative research approach to investigate existing literature and solutions as well as community based participatory research (CBPR). Each cycle includes stakeholder input in the planning, analysis, and debugging phases. CBPR stakeholders include: 1. Client Group, which includes children, teens, and families facing a diagnosis of terminal illness; 2. Provider Group, which includes healthcare and mental health care practitioners who provide care to those facing a diagnosis of terminal illness; 3. Expert Group, which includes experts in relevant research, policy, and practice.

**Results.** This poster presents the outcomes of the Planning and Prototype cycles.

**Planning Cycle.** The Planning Cycle established the need for a coding, robotics, and making based psychoeducational intervention for children, teens, and families facing a diagnosis of terminal illness. The results of a thorough review of the literature as well as what was planned by parents with cancer and their children, and planned, analyzed, and debugged by a provider group of oncological social workers yielded the following claims:

Children and teens facing familial terminal illness need information, coping skills, and a supportive community. Psychoeducational intervention can provide information, coping skills, and support when therapists and delivery are interactive and engaging. Client engagement is essential yet challenging with children, teens, and families. Play therapy leads to client engagement. Electronic games show promise in leading to client engagement. Robotics combines the proven aspects of play therapy and electronic games in engaging clients.

**Prototyping Cycle.** The Prototyping Cycle yielded a preliminary intervention model. The intervention model has been tested for design, but not content, by over 200 children, teens, and young adults who have neither a terminal illness nor live with a family member who is terminally ill known to principal researchers. This test is meant to provide a cross section of user demographics and test the viability of the model in practice. Results of these interactions will be analyzed by Provider Group participants including generalist teachers and paraprofessionals and debugged by Maker educators, STEM educators.

*Pamela Davis (Wellbotics)*

## Interface Design / Evaluation

### K-1

#### Online Timeline Followback Survey

The Online Timeline Followback Survey (O-TLFB for short) is a web application developed as a substitute for the already popularized TLFB in-person questionnaire. O-TLFB is a modular program developed in python that allows researchers to ascertain the drug and alcohol usage of participants over a specified timeline. Compared to the typical pen-and-paper version of the survey, the O-TLFB allows

participants to fill out the form from any location and submit their results in near anonymity. Once the survey has been completed the program automatically processes the data, calculates aggregate sums such as total drinking days or average amounts of a substance taken per day, and then uploads the transformed data to an online database for further analysis. This application is both easier and faster than the paper processes currently in use, and follows the current trend in research towards more online tools.

As of now this tool has been used in our lab for months now, and general trends have been analyzed.

*William Newman (Center for Innovation and Creativity)*

## Robotics

### L-1

#### Food-Delivery Robots on University Campus: Profiles of Users and Non-Users

**Statement of the problem:** Technological advancements in robotics and the growing acceptability and adoption of automation from the public have fueled a significant uptake of robotic applications in the food industry [1,2]. As restaurants develop innovative dining services, they also look for safe and cost-effective delivery services [3]. Food delivery robots were introduced to our university campus two months ago. This new service has generated significant media attention, public interest, and discussion on campus. However, objective assessment of the service, from user's and non-user's perspectives, has not been collected.

**Procedures:** An on-line survey was developed to collect social interactions and impressions of the robots, reasons and experiences for using and not using the food-delivery service in the past two months, as well as demographic information. The survey targeted university community members, including students, staff, faculty, and visitors, and has been disseminated through university-wide email lists as well as in-person recruitment at high-traffic areas on campus. The study received approval from the university Institutional Review Board.

**Analyses and Results:** Data collection is on-going and data will be analyzed prior to the conference. Results will be presented at the conference. The analytic plans include 1) comparing the demographic information (e.g., age) between users and non-users; 2) comparing the impressions (e.g., machine-like vs. human-like) of the robots between high frequent, low frequent, and non-users; 3) comparing the perceived positive and negative social interactions with robots (e.g., robots get in my way when walking around campus) between high frequent, low frequent, and non-users; and 4) comparing features (e.g., acceptable wait time) of the service between high frequent, low frequent, and non-users.

**Practical Implications:** This study is expected to lead to an objective assessment of the social interactions with the food-delivery robots as well as the perceived benefits and concerns from the context of a university campus. Results are also expected to provide guidelines for refining the implementation and marketing strategies, and to lead to the development of other delivery services (e.g., packages) that could benefit from automated robots.

**Conclusions:** As the society embraces new, innovative technologies that have the potential to enhance mobility and convenience, systematic evaluations of user experience and interactions remain an essential step in assuring safety and sustainability.

*Yi-Ching Lee (George Mason University)*

## Simulation and Modeling

### M-1

#### **Immersive Simulation Scenario-embedded Cognitive Assessment: Towards Integrated Psychometrics with Performance Probes, Biomarkers and Machine Learning**

Background: Self-report is convenient psychometric modality. Its limitations, ranging from social desirability to malingering and self-deception, have driven psychometric innovation and invention for over a century, from Binet and Raven to Rash, IRT, Computerised Adaptive Testing and most recently, Game-based assessment. Modern technologies enable the development of dynamic scenario-based, immersive measurement protocols to assess knowledge, skills, traits and abilities that don't look and feel like psychometric tests - instead offering an engaging, game-like experience to the test taker. Building measurement protocols into such realistic simulation scenarios is a potentially disruptive alternative to incrementally improving traditional psychometric formats. The resulting simulation-embedded assessment must meet stringent reliability and validity criteria. Biomarkers derived from physiological responses to cognitive-behavioral tasks offer a potent source of objective data to cross-validate the simulation-embedded metrics. Our study examined the relationships between simulation-derived measures of individual differences related to resilience and adaptability, and the theoretically-relevant biomarkers. The method builds on the emerging work in psychometric gamification, and combines it with the search for evidence-based measurement protocols utilising wearable technologies for physiological monitoring.

Method: Ninety undergraduate students volunteered to perform a simulated driving task developed to represent an emergency vehicle (Ambulance) in a time-pressured training race through a dense urban environment. The driver's task was to complete a designated route in dense city traffic following the directions given by green arrows at each intersection. The instruction presented the task as an emergency driving training course that emphasised safety but allowed the driver to break speed limits and other road rules. Drivers were exposed to five traffic hazard-type events (falling lamp posts, black ice, animals on the road, heavy fog, and road blocks) which were planted at different locations and time points within each of the five laps comprising the task. These hazard events served as probes to simulate different stressor types, and to strain attentional and cognitive resources. Participants' eye-blink rate, pupil diameter, and heart rate (HR) were continuously monitored during the simulation. Criterion constructs of resilience and adaptability were measured with traditional psychometric scales.

Results and Conclusions: Performance dynamics, including linear and non-linear trend analytics, were examined to classify participants' actions within the task. For each individual the magnitude and direction of performance changes were estimated over their baseline, including linear improvements, improvements followed by a decline (quadratic trend), and recovery (cubic trend). The results show that simulation-derived metrics can successfully predict target traits and abilities, including adaptability and resilience, using machine learning techniques and trend analyses. The data from a sub-sample (N=37) also show meaningful relationships between bio-markers and behavioural estimates of baseline performance and its sustained dynamics. When analysed across the entire duration of the driving task, blinking rates predicted speed-related

performance metrics, while pupil diameter slope metrics predicted collisions. When restricted to the time windows around the traffic hazard probes, the dynamics of pupil diameter (linear slopes) correlated negatively with the increase in collisions during the probes ( $r = -.34$ ), and positively with the baseline collisions count and its recovery dynamics ( $r$ 's = .53 and .65). Similar slopes representing the HR dynamics correlated with the lowest collision count during the probe events ( $r = .46$ ). Overall, the results demonstrate a promising convergence between the behavioural and bio-markers of performance in our simulated driving task. While our validation criteria have so far been limited to self-report variables, they will ultimately be replaced with real-world health and performance outcomes. And when such real-world validation succeeds, it may render the self-report metrics redundant. The avenues to extend the generalisability of this new biomarker-enhanced, simulation-embedded assessment methodology are discussed, along with its potential applications in high-stake assessment.

*Eugene Aidman (University of Sydney / DST)*

## Social Cognition

### N-1

#### **Evaluating the Substantive Aspect of Construct Validity for Online Social Cognitive Processes**

INTRODUCTION: A growing body of experimental and longitudinal evidence reveals that aggressive children and adolescents have deficits in social information processing (SIP) responses, including (a) encoding, (b) interpretation, (c) goal selection, (d) creating possible responses, and (e) making decision (Crick & Dodge, 1996; Dodge, Coie, & Lynam, 2006; Lansford, Malone, Dodge, Crozier, Pettit, & Bates, 2006).

The increased use of technology among adolescents has allowed students to apply SIP skills in online environments. To date, no study has examined SIP skills in online environments and how these skills are linked to online fights (see Dooley, Pyżalski, & Cross, 2009). As decision making is one of the important skills in SIP responses, the purpose of the current study was to develop and explore the substantive aspects of the construct validity of a self-report measure for the decision-making skills of Social Information Processing in online environments (OSIP-DE), particularly, when adolescents experiencing online fights. OSIP-DE is a modified version of Kupersmidt, Stelter and Dodge's (2011) Social Information Processing Application scale (SIP-AP) which is designed only for boys (age 9 to 12) to response to a negative outcome for the victim and the purpose of the perpetrator in online settings, using Vignette.

Statement of Problem: This study addresses this gap by investigating two research questions: 1) What strategies do adolescents apply when responding to the OSIP-DE items? 2) What are the general content topics that come up for adolescents when responding to the OSIP-DE items?

METHODS: Participants: 21 adolescents (76% girls) in grade 8-12 from a secondary school in West Vancouver, the Lower Mainland of British Columbia. Methods: 1) Think-aloud protocols as social cognitive interviewing techniques; 2) Self-report demographic background and OSIP-DE (6 items) were used to assess adolescents' responses of their own decision when they are in an online fight across a variety of situations.

RESULTS: OSIP items are displayed in Table 1. As shown in Figure 1, adolescents used the absolute and relative strategies as well as the following content categories for all items, including 'values', 'emotions', 'personal characteristics', 'consequences', 'coping strategies', 'relative situation', and 'relative coping strategies'.

There were also significant differences between absolute and relative strategies. The absolute strategy being used more often than relative strategy in item 1 (chi-square (1, N=21) = 17.45;  $p < .001$ ), item 2 (chi-square (1, N=21) = 17.45;  $p < .001$ ), item 3 (chi-square (1, N=21) = 16.1;  $p < .001$ ), item 4 (chi-square (1, N=21) = 17.45;  $p < .001$ ), item 5 (chi-square (1, N=21) = 17.45;  $p < .001$ ), and item 6 (chi-square (1, N=21) = 17.45;  $p < .001$ ).

CONCLUSION: This study provides insights into the strategies that adolescents used to respond to OSIP-DE. The findings showed: 1) Think aloud protocol analysis can be considered as a method to determine strategies used by teenagers on experiencing online fights; 2) The adolescents' item responses are governed by strategies that are meaningful and reflect ideas on the cyber-aggression literature; 3) The majority of the teenagers did not have problems understanding the OSIP-DE items. However, further research is needed to replicate these findings with teenagers of diverse age groups and different ethnic backgrounds when responding to the items of the OSIP-DE.

*Rose Maghsoudi (University of British Columbia)*

## Social Networks

### O-1

#### **Psychological Distress and the Use of an App for Social Support Among Men and Women with Fertility Concerns**

Background: Feelings of isolation are common among infertile people (Greil, Slauson-Blevins, & McQuillan, 2010; Hinton, Kurinczuk, & Ziebland, 2010). Online forums and chat rooms allow infertile people to exchange stories and experiences and to receive empathic reactions from others who understand the medical and emotional challenges of infertility (S. H. Malik & N. S. Coulson, 2010). Despite these advantages, there are a number of problems associated with online support groups. Users express concerns about misinformation (Malik & Coulson, 2008), and lack of responses to one's own comments and questions (S. Malik & N. S. Coulson, 2010). When asked to indicate preferred features of online peer support, men and women seeking fertility treatment highlighted mobile accessibility and monitored peer-to-peer communication (Grunberg, Dennis, Da Costa, & Zerkowitz, 2018). The present study aimed to evaluate an app designed to provide medical information and psychosocial resources for men and women undergoing fertility treatment. The app included a message board monitored by trained and supervised peer supporters. The proposed presentation examines indices of mental well-being in 3 groups of app users: active posters on the board, lurkers who read messages but did not post, and people who did not access the message board. We also evaluated message board use via responses to open-ended questions about user preferences. Methods. A sample of 159 women and 42 men were recruited at fertility clinics in Montreal and Toronto, Canada to test the Infertility app, which was developed by researchers and fertility health care providers to provide information on the medical and psychological aspects of fertility treatment, as well as a message board where users

could discuss issues of concern to them. Volunteers with experience of infertility and its treatment each provided 5 hours/week to monitor the board, participating in discussions and ensuring that every user who posted a comment or question received a response. These peer supporters were trained and supervised by coordinators, including a doctoral student in clinical psychology, a social worker, and a couple and family therapist. Upon recruitment, study participants completed questionnaires, including the 4 item short form of the Perceived Stress Scale (PSS-4) (Warttig, Forshaw, South, & White, 2013), and the Generalized Anxiety Disorder 7-Item Scale (GAD-7) (Kroenke, Spitzer, Williams, Monahan, & Lowe, 2007). The PSS-4 is a self-report measure of the amount of stress encountered in daily life during the past month, and the ability to handle the stress, with the 4 items rated on a Likert scale from 0 (never) to 4 (very often). The GAD-7 assesses symptoms of anxiety in the past two weeks, with items rated from 0 (not at all) to 3 (nearly every day). After 6 weeks of app use, participants were asked to describe which features of the app they liked best, and which they liked least. Google Analytics were used to determine whether participants used the message board. Based on these data, participants were classified as active posters, lurkers, or non-users of the message board. One way analysis of variance was used to compare these 3 groups of app users in terms of their scores on the PSS-4 and the GAD-7. Responses to the open-ended questions about message board use were compiled and examined for general themes. Results. Of the 201 app users, 37 (18.4%) were active posters, 61 (30.3%) were lurkers, and 103 (51.2%) did not use the message board. The vast majority of active posters (35/37) were women. Men comprised 18% (11/61) of the lurkers. Compared to lurkers, active posters reported significantly higher levels of perceived stress and more anxious symptomatology (see Table 1). The difference between posters and non-users approached significance, but there was no difference between lurkers and non-users in stress or anxiety. In response to the open-ended questions, many participants noted that the message board was the best feature of the app. The board provided "someone to vent to", and helped users to calm down. Users appreciated reading about the experiences of others, and having the opportunity to share their own feelings and concerns. Lurkers indicated that even if they didn't post, they were comforted by the support given to others. Peer supporters were seen as being available, and ensured that the board remained respectful, allaying concerns about other forums on the internet. Discussion. The Infertility app message board was designed to provide social support to people undergoing fertility treatment, in accordance with user preferences regarding mobile accessibility and monitoring. This feature of the app was of particular interest to users experiencing psychological distress. The monitored message board seemed safe and reliable, providing opportunities to share one's own experiences, and derive comfort and useful information from the experiences of others. Future research should examine whether this type of intervention may serve to improve mental well-being and quality of life among people undergoing fertility treatment.

*Phyllis Zerkowitz (Jewish General Hospital and McGill University)*

## Theoretical Frameworks

### U-1

#### Technology, Culture, and Society Initiative: On Bringing a University Together

Computer scientists think often of Ender's Game these days. In this award-winning 1985 science-fiction novel by Orson Scott Card, Ender is being trained at Battle School, an institution designed to transform young children into military commanders fighting against an unspecified enemy. Ender's team engages in a series of computer-simulated battles, eventually destroying the enemy's planet, only to learn then that the battles were very real and a real planet has been destroyed.

In a similar vein, many computer scientists got involved in computing because programming was fun at its essence, and the benefits of computing seemed intuitive and obvious. Like Ender, however, we are realizing that computing is not a game or utopia—it is real—where one should anticipate real costs along with any real benefits. Social scientists are therefore critical in partnering with computer scientists, to understand and contend most effectively with real-world problems where technology and society are intertwined.

Consider three examples. First, much has been written recently about the adverse impact that automation has had on working-class people—an impact that has already had profound political consequences—with even greater impact expected as fundamental tasks and jobs get automated, such as individual and commercial driving, or checking out at a department store. Second, it has now become painfully clear that “frictionless sharing” on social media has given rise to the fake-news phenomenon. It is now widely accepted that this has had serious impact on both the 2016 Brexit referendum in the UK and the 2016 U.S. presidential election. Third and finally, a 2017 paper in *Clinical Psychological Science* attributes the recent rise in teen depression, suicide attempts, and suicide to the ascendance of the smartphone.

We discovered fire about one million years ago, discovery so crucial to human progress and development that Greek mythology attributes it to Prometheus stealing it from the gods to give it to humans, but people still die from fire regularly. Likewise, technology has always been a double-edge sword. History has shown that human civilizations have never been able to stop technological advancements. Instead, we learn over time what technology's adverse consequences are on society, and how to deal with them. Much like an uncontrolled fire, uncontrolled information technology poses another major challenge to the developed world, which is the erosion of shared prosperity. Without shared prosperity, societies polarize and liberal democracy declines, as we have witnessed recently. The negative consequences of technology on society are significant, and they will likely remain just as profound in the future; yet very little academic research and scholarship systematically addresses these technological and societal phenomena that give rise to these consequences, or how to prevent or intervene on these consequences.

The poster we present reflects our proposal at Rice University: To kick-start a university-wide initiative focused on the intersection of technology, culture, and society. The goal is to create a permanent intellectual and multidisciplinary center at Rice University dedicated to the descriptive and prescriptive study of a sociotechnical system—namely, how technology impacts society and culture, and how society responds to such impacts. This center will engage in three types of activities: (1) Research and scholarship: promoting interdisciplinary research studying sociotechnical challenges and policies,

such as: economic impact of technology, the importance of work to well-being, technology and urban development, and public policy for cybersecurity, and (2) Education: developing undergraduate and graduate courses at Rice focusing societal and ethical considerations of technology, and (3) Outreach: organizing lectures, workshops, and conferences aimed at both the broad academic audience and to the general public.

Towards the establishment of such a center, we have already established an advisory board that meets regularly. This board consists of faculty members from all academic schools at Rice University (engineering, social sciences, humanities, and natural sciences, as well as two think tanks (Baker Institute and Kinder Institute). Development efforts are well underway.

We hope that you will find this poster to be a compelling and inspirational example of how departments across the university—in the social sciences and beyond—can collaborate to serve one another and society writ large.

*Fred Oswald (Rice University)*

## Trust / Control

### P-1

#### Insights into Trust in Everyday Technologies

Over the last several decades, technology has increased in all aspects of our daily lives, from on-the-job, to at-home, to leisure activities, to doctor visits. In fact, one could hardly function appropriately in today's world without access to technological advances ranging from email to modern washing machines. Indeed, technology has become essential even in our day-to-day interactions with others. It has become crucial to understand how humans interact with, and rely on, these technologies. A user's trust in the technology is a key component of understanding human-technology interaction, but little is known about ratings of trust for everyday technologies. There has been considerable research into trust in automation, demonstrating the key role such trust plays in use of the technology. Both initial adoption and use of technologies is impacted by trust (Parasuraman & Riley, 1997). After establishing initial trust for adoption, then trust calibration is needed to prevent over-trust, which can lead to overreliance (Lee & See, 2004; Parasuraman, Sheridan, & Wickens, 2008). Trust is complex and has many components. Prior research has demonstrated that trust can be broken into distinct components: trust and distrust (Jian, Bisantz, & Drury, 1998).

Just as with trust related to automation, it is important to also understand trust in relation to systems and devices that are not so far along the “high tech” dimension. Thus, the goal of this study was to explore trust people have for everyday technologies. To evaluate trust and distrust in 23 technologies, 193 participants were recruited from a university population, or via Amazon Mechanical Turk, to complete an online survey. Of those, 15 were excluded because they failed at least one of the attention checks embedded in the survey. 178 participants were included in the final data set, with a mean age 33 years (SD=14.4, range: 18-72). There were 84 (47.2%) males and 94 (52.8%) females.

Using a trust-in-automation scale developed by Jian et al. (1998), participants reported trust for each of 23 technologies such as: camera; dishwasher; e-book reader (e.g., Kindle); electric can opener; electric razor; and wearable health technology (e.g., Fitbit).

Overall, participants reported moderate levels of trust for all the 23 technologies ( $M=5.02$ ,  $SD=1.28$ ), with some interesting differences. The highest level of trust was for safety glasses ( $M=5.84$ ,  $SD=1.08$ ) and wheelchairs ( $M=5.66$ ,  $SD=1.18$ ). The lowest level of trust was for intelligent personal assistants (e.g., Alexa) ( $M=4.10$ ,  $SD=1.45$ ) and smart speakers (e.g., Google home) ( $M=4.27$ ,  $SD=1.47$ ). Correspondingly, participants on average reported low distrust for the 23 technologies ( $M=1.80$ ,  $SD=1.08$ ). The highest level of distrust was for mobile phone apps ( $M=3.22$ ,  $SD=1.54$ ) and intelligent personal assistants (e.g., Alexa) ( $M=3.02$ ,  $SD=1.69$ ). The lowest level of distrust was for safety glasses ( $M=1.25$ ,  $SD=0.71$ ) and washing machines ( $M=1.28$ ,  $SD=0.76$ ).

The moderate levels of trust, and limited distrust, make sense given that the technologies are commonly used among the overall population. Technologies rated highest for trust, and correspondingly lowest for distrust, were those that have been designed for safety (safety glasses), as an assistive technology (wheelchair), or for basic household tasks (washing machine). In contrast, the technologies that were rated lowest for trust, and correspondingly highest for distrust, were relatively newer technologies, with which people may not have as much familiarity, and for which people may not have as solid a mental model.

This study demonstrates that a moderate level of trust may be all that is needed for use for everyday technologies, but there is still some variability among them that needs further exploration. The distinction between trust and distrust is important to recognize in both research and product design. Our ongoing next phase of research is exploring how trust and distrust is related to technology adoption and frequency of use; and is exploring an even more extensive list of newer technologies.

*Rachel Stuck (Georgia Institute of Technology)*

## P-2

### Calibrating Trust Towards an Autonomous Image Classifier in an Image Classification Task

As technology continues to advance, our computers become more intelligent, increasing their capacity to operate autonomously on complex tasks. For these artificial intelligence systems to be successful, individuals collaborating with them need to be able to trust them in order to use them correctly. Ideally, user's trust towards a system will accurately reflect the machine's actual performance. Incorrectly calibrated trust is problematic, too little trust in a system can lead to disuse, whilst too much trust can lead to misuse and overreliance (Parasuraman and Riley, 1997). The most consistent influencer of trust towards autonomous systems appears to be task performance (Chavailleaz, Wastell, and Sauer, 2016; Hoff and Bashir, 2015). If a system commits obvious errors or behaves erratically, trust towards it can diminish rapidly, and can be difficult to recover (McGuirl and Sarter, 2006; Merritt, Heimbaugh, LaChapell, and Lee, 2013). Thus, for successful human-machine collaboration, the user's understanding of machine performance is critical, yet evaluations can be difficult when task performance is ambiguous. When performance is unclear, users' evaluations may be more susceptible to influence from implicit attitudes and biases towards automation and technology (Merritt et al, 2013; Goddard, Roudsari, and Wyatt 2011). For example, individuals that score higher on criteria for Perfect Automation Schema, the belief that automation performance is almost flawless, are more

likely to display intense decreases in trust following errors, meaning their evaluations do not accurately represent machine performance (Dzindolet, Pierce, Beck, and Dawe, 2002; Merritt, Unnerstall, Lee, and Huber, 2015).

However, users' trust can be improved by including a metric of machine confidence within the interface, which illustrates the system's calculated probability in its decision being correct. This has been implemented across a range of autonomous system types, such as robots and flight simulators. Furthermore, confidence itself has been represented in a variety of ways, some simple, such as High, Medium, Low confidence categories, and some complex, such as visual representations using objects that vary in colour, shape and opacity (Kaniarasu, Steinfeld, Desai, and Yanco, 2013; Selkowitz, Larios, Lakhmani, and Chen 2017) Across multiple studies, confidence metrics appear to help calibrate user trust more accurately to autonomous systems, even when errors occur. Yet, in some cases decision support information that is overly complex can be detrimental to users' trust, owing to information overload (Selkowitz et al, 2017). However, within these studies it is unclear just how transparent machine performance was to participants, often it appears participants were aware that performance was either good or bad, suggesting a lack of ambiguity in performance. If performance is clearly poor yet machine confidence is low, it is easy to imagine users will have more trust towards a system, or vice versa. However, it is unclear how confidence metrics will influence trust when performance appears ambiguous and is difficult for users to evaluate. We sought to explore this, by investigating participants evaluations of an autonomous image classifier, which varies in error and success rates, on a series of images that vary in their clarity and identifiability to human users.

Participants completed an image classification task, and evaluated the AlexNet image classifier assisting them (Krizhevsky, Sutskever, and Hinton, 2012). They attempted to correctly label a series of objects within images, and interacted with the classifier through a graphical user interface. Across four conditions of this task, within each trial we examined the Dependent Variables of: trust towards the classifier; perceived machine classification accuracy; and participants' familiarity with each image. These factors were examined across a variety of trials in which classifier performance ranged from: unambiguously good; ambiguously good; ambiguously poor; and unambiguously poor. To explore evaluations of classifier performance, participants used four different interface versions, which varied in their complexity of displaying the classifier's task confidence. This was represented as one of the following: a graphical representation of confidence being low, medium or high; confidence as a numerical percentage (0-100%); or visually using a graph that illustrated the classifier's top 5 closest predicted image label probabilities; whilst a control interface was also included without any decision support information. Across these versions, interface designs were kept visually similar to limit the influence of aesthetics. After using each interface version, participants completed the NASA Task Load Index, to measure factors such as ease of use and task demand (Hart and Staveland, 1988). Participants also completed the Propensity to Trust Machines questionnaire (Merritt, 2011), which allowed for consideration of participants general tendencies to trust machines. Collectively, this will allow us to develop our understanding of the roles of perceived performance, support information, and trust, across both ambiguous and unambiguous performance. Data collection from 60 participants will conclude by August 30th, and

will be analysed using linear mixed effect models. This will investigate the roles of perceived performance, ambiguity, user attitudes and decision support information, when predicting trust within human-machine classifying teams.

*Martin Ingram (University of Glasgow)*

## Video / Computer Games

### Q-1

#### Feeling and Working in Digital Appalachia: Two Months Playing Fallout 76 Fosters a Sense of Place in West Virginia

**Problem Statement:** “Someday, people will visit ideas instead of places” - Toba Beta. Bethesda Softworks’ Fallout 76 was among the most highly anticipated video game releases in recent memory (Farokhmanesh, 2018). As with previous franchise titles set in Washington D.C. (Fallout 3) and Boston (Fallout 4), Fallout 76 (FO76) is situated within the physical geography, authentic locations, and local folklore of West Virginia. FO76 developers partnered with the WV Tourism Board to “[show] the rest of the world what a gem West Virginia is” (WV Governor Jim Justice, as cited by Griffith, 2018; para. 3).

While most mediated portrayals of Appalachia engage pejorative stereotypes (Bowman & Groskopf, 2010), FO76’s design can be viewed through the lens of Tuan’s (1975, 1979) cultural geographic notion of a sense of place in a more neutral and self-referential light. For Tuan, experiences of spaces go beyond mere recognition of labels to include senses of subjective knowing—“the perception of meaning associated with a site or location” (Bott, 2000, p. 7). Likewise, advances in interactive media allow users to engage perceptually sprawling online locations that could hold meaning for their users (Possler, Klimmt, & Raney, 2018), which calls to question if individuals engaging digital spaces could develop a sense of place for the analogous real-world locations. Furthermore, Bowman (2018) suggests that gaming outcomes (such as sense of place) can be understood as a function of how demanding the digital experience is of the player’s cognitive, emotional, physical (control and exertion), and social resources.

Given the paucity of empirical work on the nature and dynamics of sense of place in digital spaces, we leverage FO76’s reflection of WV as a physical and cultural place to ask: (RQ1) how do video game demands influence the formation of sense of place?

**Procedures:** A subset of FO76 players (N = 369) from a larger study on player experiences completed three online surveys: two weeks prior to game’s release (T1), two weeks following release (T2), and two months following release (T3). This sample was 61% female, age M = 31.82 (SD = 8.72), and politically neutral, M = 3.65 (SD = 1.82; scale 1-7, liberal to conservative). About 24% were WV natives; 61% could name a real WV location prior to release (with 86% at T2, 87.8% at T3). Respondents played FO76 for M = 116.08 total hours (SD = 89.13) and rated the game experience as moderate (M = 75.76; SD = 17.23, scale of 0-100).

Sense of place was measured using a modified version of Bott (2000). The suggested five-factor measurement model did not fit observed data (see supplemental analysis at <http://bit.ly/2BYKheC>), was reassessed using principal axis factoring with oblique rotations. A unidimensional factor structure was suggested, with 13 items

having factor loadings > .600 and explaining 60% of variance: T1 M = 4.74, SD = 1.25,  $\alpha = .949$ ; T2 M = 5.07, SD = 1.17,  $\alpha = .945$ ; T3 M = 5.08, SD = 1.19,  $\alpha = .955$ . Perceived demands were assessed using Bowman, Wasserman, and Banks’ (2018) seven-point Likert scales for cognitive (M = 3.64, SD = 1.30,  $\alpha = .844$ ), emotional (M = 3.98, SD = 1.61,  $\alpha = .928$ ), controller (M = 5.69, SD = 1.19,  $\alpha = .858$ ), exertional (M = 2.61, SD = 1.48,  $\alpha = .928$ ), and social demand (M = 4.39, SD = 1.30,  $\alpha = .854$ ).

**Analysis and Results:** A multiple regression was conducted regressing T3 sense of place scores onto participant demographics and pre-gameplay (T1) sense of place and WV location recall (Model 1), total gameplay hours, overall game rating, and T2 WV location recall (Model 2), and the five demand factors (Model 3). The overall model was significant (Table 1) and explained > 65% of variance, adjusted R<sup>2</sup> = .651, F(5,352) = 11.08, p < .001, Durbin-Watson = 2.27. Significant control effects of T1sense of place ( $\beta = .607$ , p < .001), political orientation ( $\beta = .081$ , p < .001), and T3 recognition of WV places ( $\beta = .274$ , p = .019) were found. Focal to RQ1, increased emotional demand ( $\beta = .269$ , p < .001) boosted and increased exertional demand ( $\beta = -.121$ , p = .006) decreased perceptions of sense of place.

**Implications and Conclusions:** Players who felt a higher emotional pull into and less exertion from FO76 were more likely to connect with West Virginia as a self-referential and meaningful place. Replications of this work might consider immersive applications beyond games, such as immersive journalism (Pavlik & Bridges, 2013), cause-related storytelling (AUTHORS), and virtual tourism (Tussvadiiah, Wang, Jung, & Dieck, 2018). Digital spaces can foster an authentic sense of place for their physical counterparts, so long as they are emotionally gripping and not too strenuous.

*Nick Bowman (Texas Tech University)*

### Q-2

#### Leading Parent’s Role in Co-Use of Video Games

**Introduction:** It is widely recognized that family is an important resource for learning media literacy [1,2, 4], but family structure is explored without distinguishing the roles of father and mother and their strategies of co-usage of media in children. Russian fathers remain the main mediators of children’s inclusion in use of media [1]. Nikken and Jansz in their research of video gaming revealed early that parents distinguished three types of strategies of mediation: (1) ‘restrictive mediation’, (2) ‘active mediation’, and (3) ‘co-playing’ [5]. Researchers recognize several types of families according to following social and psychological criteria: closeness-openness of the family circle, distribution of power-responsibility, degrees of emotional proximity among family members [6].

The main hypothesis of our investigation was that distribution of power-responsibility in family determinates leading parent’s strategy of co-use of videogaming in children.

**Method:** Sample of our research consisted of 30 dyads—teenagers in the age of 10-18 years old (17 boys and 13 girls) years and have both parents living together. According to the information about their income all participants belong to middle class. Most of the fathers (23 out of 30) have university degrees, while the rest obtained only secondary education. We used the following methods: 1) in-depth interviews with children and their fathers separately on how much time they spend together using computers, how often and who makes decision when to start and to stop playing video games, how

fathers react on children video gaming, etc. 2) our original graphic method “Family models”. Method of “Family model” is a technique designed to identify the distribution of power and responsibility between family members. In our study we offered children to choose out of five models of family in order to identify the role of father in family structure.

Results and discussion: The groups of fathers had been mapped out according to their position within family structure and mediation strategies. The first group (eight out of thirty fathers) includes strictly controlling fathers; they consider video gaming as a waste of time, with no benefit for children development and self-esteem. The second group - “qualified” fathers (six out of thirty) uses the reflexive strategy of modeling online activities in children; to teach children how to regulate media use (more conscious use, play, more conscious approach to media overall). The third group of fathers (16 out of thirty) uses motivating mediation strategy—co-playing; they prefer to play video games themselves, increase the enthusiasm of children. Fathers that are main power holders (dominant fathers) tend to control the formal aspects of computer usage by children—the time of play, readiness of a child to submit to the imposed schedule of usage. Fathers from families based on partnership tend to participate in choosing, planning and joint video gaming for pedagogical goals. They could be described as active mediators. Submissive fathers who tend to give up the power in favor of their wives occupy “child’s position” in family and play video-games themselves. Among all the fathers we observed the tendency to more co-use video games with their sons. This gender selectiveness is explicit even in case with dominant controlling fathers, since videogaming is “something that boys do”.

Conclusion: We can regard family structure as an important predictor of strategy that fathers choose to mediate the co-use of video games with their children. Development of new digital products and video games should account for role of leading parents and their readiness for co-use of media. This data can be used in designing a new generation of video games, as well as training programs for mentors and parents.

*Olga Makhovskaya (Russian Academy of Sciences Institute of Psychology)*

### Q-3

#### **Mental Health Representation in Games: State of the Research**

Since 2014, there has been a dramatic increase in games research and media coverage centered on the representation of diverse populations and experiences in digital games. These discussions have focused on the lack of non-white, non-male characters and on stereotyped portrayals of gender and race (Jenkins & Cassell, 2008; Williams, Martins, Consalvo, & Ivory, 2009; Shaw, 2014; Sarkeesian, 2013; Salen & Zimmerman, 2003). As this area of study continues to develop, it is important to expand research and analysis of representation to include less-visible marginalized groups, such as persons with mental illnesses. Approximately 24% of video games portray mentally ill characters (Shapiro & Rotter, 2016) and the presence of mental illness-related content, such as insane asylums and straight jackets, or pejorative terms like “crazy” and “psycho”, increases the frequency with which players come in contact with mental health representations. Compared to other marginalized groups, the portrayal

of mental illness is common, yet research on the portrayal of mental illness in digital games is practically non-existent (Ma, 2017).

How games portray persons with mental illnesses and psychopathological content conveys a sense of their cultural value (Klein & Shiffman, 2009) and, conversely, our cultural values take cues from the media we consume (Lule, 2016). Over 40 years of media research has consistently found that mass media is the most common public source of information about mental illness and that media portrayals of mental illness tend to be negative, exaggerated, and inaccurate (Singorielli, 1989; Wahl, 1992; Stout, Villegas, & Jennings, 2004; Ma, 2017). Exposure to negative media portrayals of mental illness have been linked to negative or stereotyped perceptions of the mentally ill among people without a mental illness (Stout, Villegas, & Jennings, 2004; McGinty, Webster, & Barry, 2013; Ma, 2017) and impaired help-seeking behaviors and treatment adherence for persons coping with a mental illness (Stuart, 2006; Maier et al., 2014).

Before we can systematically analyze how mental illness is portrayed in games, we first need to identify what constitutes mental illness in games. What does it mean for a game to portray mental illness? This paper will examine previous attempts to identify and analyze mental illness in video games, create an updated framework for categorizing mental illness portrayals, and provide examples of classic and contemporary games that fit those categories.

*Kelli Dunlap (iThrive Games)*

## Virtual Environments / Augmented Reality

### R-1

#### **Current and Future Applications of Augmented Reality in Higher Education**

As augmented reality (AR), or the layering of the digital into the real world, is becoming a larger part of society, it has also become part of the here and now of education more generally—and not just higher education. Students in K12 environments are entering higher education having been exposed to augmented reality in a range of ways (see, for example, Cai, Wang & Chiang, 2014; Hsu, 2017; Huang, Chen & Chou, 2016; Radu, 2014). Indeed, as the tools become easier to use, instructors at all levels are now creating their own augmented reality experiences for their students (Crews, 2018).

As the use of augmented reality increases in K12 classrooms, students may begin to expect these same technological innovations in higher education as well. Although the literature review by Akçayır and Akçayır (2017) found that over half of research studies on AR focused on K12 populations with only 29% of studies focusing on university students and seven percent focusing on adult learners, we can expect studies of augmented reality in higher education to increase as the technology evolves and becomes less expensive to implement. Current studies of the use of augmented reality in higher education, similar to their K12 study counterparts, have also shown benefits and challenges for this tool. Some of the benefits include decreased time for students to learn new material (Rizov & Rizova, 2015) and increased development of spatial thinking (Carbonell Carrera & Bermejo Asensio, 2017).

At Oregon State University Ecampus, we have begun experimenting with the development of AR in a range of classroom environments. For example, one project titled the “Dog Skull Augmented

Reality Interactive” allowed for an instructor to digitize borrowed materials so that a larger group of students could have access to viewing and manipulating 3D images of canine skulls. For this project, one of the co-presenters used Vuforia, a marker-based augmented reality framework. By linking marker graphics to models via the Vuforia app, users can teach devices (such as iPads or other mobile devices) to recognize the marker and place the model into a camera overlay. The camera uses the size, orientation, and skew of the marker to position the model in 3D space where students can interact with the 3D image through the use of an app. Students using this technology now have access to course materials that would otherwise be unavailable to them.

Based on the presenters’ experience with integrating AR into online and hybrid classroom environments, this poster presentation offers an overview of the benefits and challenges of applying AR within higher education contexts. Current examples of augmented reality use will be offered to demonstrate how AR can enhance teaching and learning in multiple fields in higher education. Additionally, through the “Dog Skull Augmented Reality Interactive” example, conference participants will be able to try a form of augmented reality as they view this poster. An additional supplementary handout will offer tips and strategies for implementing augmented reality in higher education.

Although augmented reality may seem too futuristic for the current higher education classroom, the technology and applications that support AR are developing at a rapid rate. While little evidence suggests that younger students will show greater ability to learn technology (i.e. “digital natives” Prensky, 2001), students who experience AR in their early educational experiences may become disillusioned if higher education does not follow suit. Just as many other technologies that we interact with in our day-to-day lives have found their way into the higher education environment, we should also plan to see augmented reality appear more frequently in our classrooms in the coming years.

*Rebecca Thomas (Oregon State University)*

## R-2

### Investigating the Impact of an At-Home Virtual Reality (VR)-Based Intervention for Chronic Pain

Research within hospital settings has shown that brief virtual reality (VR)-based interventions are effective as adjuncts to pharmacotherapy or as standalone interventions for acute pain, and some evidence suggests that they can provide relief for chronic pain. To date, only a few studies have investigated VR for chronic pain management and the data are preliminary. Pilot studies on the use of VR for complex regional pain syndrome, chronic neck pain, chronic headache/migraine pain, chronic neuropathic pain, and fibromyalgia have been conducted and show promise for the use of VR to reduce pain and prolong treatment effects. However, most studies have used non-standardized techniques and tools in small-scale pilot studies. Further investigations with larger, representative and appropriately powered samples and refined methodologies must be conducted to establish clinical efficacy of this VR therapy. In addition, future studies require a suitable control environment to contrast with the VR treatment. Much of the existing VR research has taken place in lab or clinic settings. These environments can be optimized for VR systems. However, much remains to be known about the effects of regular and prolonged VR use for real-world and home applications, where

they will be used for many chronic conditions. The present study is a longitudinal randomized controlled trial examining the effectiveness of a 21-day, home-based, VR intervention in reducing chronic pain and improving quality of life. Participants are males and females aged 18-65, meeting criteria for low back pain or fibromyalgia for at least six months, and an average self-reported pain intensity score of 4 on the 11-point Numerical Rating Scale. Participants will be randomized to the VR-based treatment, which contains psychoeducation, breathing training, guided mindfulness exercises and distraction, or the audio-based control, containing similar content. Primary outcomes will include self-reported quality of life, pain intensity and pain interference via the Defense and Veterans Pain Rating Scale measured pre-intervention, during intervention, post-intervention, and at a two-week-follow-up. Analysis of within-treatment changes will be conducted using paired t-test comparisons and between-treatment differences using analysis of covariance. Current treatments for chronic pain primarily consist of pharmacological methods including opioids, interventional techniques, physical rehabilitation, and intensive psychotherapeutic programs. Mechanisms underlying chronic pain include a complex interaction of physiological, emotional, cognitive, social and environmental factors and interdisciplinary treatment is not readily available to many suffering with chronic pain. VR interventions have the potential to serve as a cost-effective, accessible, non-invasive, alternative method for pain management that reduces risks of side effects and/or opioid dependency. Data collection is currently in progress and analysis will be completed by May 2019.

*Tyler Chavez (Johns Hopkins University)*

## R-3

### Using Virtual Reality to Teach Urologic Surgical Techniques for Rare Conditions

Children’s Hospital Los Angeles (CHLA) is one of the top institutions for pediatric urology care in the United States, providing high-quality care to children with rare, complex urologic conditions, such as bladder exstrophy. Bladder exstrophy, a condition in which the bladder forms outside the body, is seen in just 2.15 babies per 100,000 live births (Nelson et al. 2005). If inadequately treated, the condition can have a significant and negative impact on the patient’s health, social life, family life, and mental wellbeing (Wilson et al. 2007; Wiersma 2008). Currently, treatment for bladder exstrophy in much of the world remains insufficient, as many patients remain untreated or receive inadequate surgical care. In Ethiopia, for example, there is an estimated backlog of 575 patients with untreated bladder exstrophy (Yousef et al. 2018). At a hospital in Durban, South Africa, 42% of the 57 babies born with bladder exstrophy from 1983 to 2006 died in early childhood due to sepsis after surgery, congenital comorbidities, or lack of medical treatment (Wiersma, 2008). Even in high-income countries, many bladder exstrophy patients fail to receive appropriate treatment due to a lack of surgeons skilled in exstrophy repair (Borer et al. 2015; Cruz et al. 2016). In China, bladder exstrophy patients have traditionally had little access to specialized surgery, and some have remained untreated through adulthood (Zhang et al. 2012.; Xiong et al. 2015). Given the rarity of bladder exstrophy, many pediatric urologists worldwide do not have the opportunity to receive in-person training on the surgical repair of exstrophy, which is considered one of the most technically challenging surgeries in pediatric urology (Cruz et al. 2016).

Observation of live surgery has been identified as a successful training method for doctors in countries across the world (Burton 2016). However, due to space limitations in the operation room (OR), typically only three surgeons are able to simultaneously observe a live surgery. Virtual reality (VR) is a promising tool that can address this limitation and enables the training of multiple surgeons at once. Every year CHLA hosts the Annual Chinese Pediatric Urology Leadership Training Course where 14 of China's top pediatric urologists come to CHLA to learn about surgical management of rare, complex conditions. In October 2018, the Division of Urology at CHLA introduced VR to the training course. Visiting urologists wore Epicenter VR Inc. headsets to observe Dr. Andy Chang and Dr. Evalynn Vasquez performing a live bladder exstrophy repair on a 7-month old girl. A urologist from Boston Children's Hospital with expertise in bladder exstrophy closure was also involved in the VR livestream via video conference, offering surgical suggestions. While observing the live surgery, visiting urologists simultaneously viewed a PowerPoint presentation, which provided an overview of the patient's medical history and the surgical techniques. With the introduction of VR, multiple surgeons can receive a 360-degree livestream of urologic surgery, with layered information enhancing the learning environment (Figures 1 & 2).

Incorporating virtual and augmented reality (AR) into urologic surgery training is safe, feasible and useful (Borgmann et al. 2017). The key aim of surgical training with VR is that enhanced knowledge is sustained, brought back to surgeons' home facilities, and imparted to future patients (Burton 2016). Using VR to livestream an exstrophy surgery, visiting pediatric urologists learned how to treat this complex condition, and brought this knowledge back to benefit their patients in China. The Division of Urology at CHLA, a Center of Excellence for Bladder Exstrophy, is committed to providing excellent care to children with bladder exstrophy, and to training surgeons across the world in successfully treating exstrophy patients. VR provides a new tool for CHLA urologists to achieve this goal. Surgeons and researchers in the Division of Urology at CHLA plan to continue exploring the role that new technologies can play in improving patient outcomes.

Zoë Baker (*Children's Hospital Los Angeles*)

## R-4

### Virtual or Reality? Same Effects of Short-Term Relaxation Scenarios on Affect and Stress

**Introduction:** The sociotechnical system approach proposes an interdependence of technical and human systems. Developments of the one require adaptations in the corresponding domain (Mumford, 2000). Digitization has a deep impact on the way of working (Parviainen, Tihinen, Kääriäinen, & Teppola, 2017). As a result cognitive strain increases which also affects productivity and performance as the latest health report of the BKK, a German health insurance, demonstrates. Employees report negative emotional and cognitive effects leading to mental stress (Knieps & Pfaff, 2017). Reversely, recovery experiences positively relate to on-job behavior (Sonnentag, 2003). Studies show that short relaxation phases affect mood positively (Kaida, Takahashi, & Otsuka, 2007) and memory performance (Lahl, Wispel, Willigens, & Pietrowsky, 2008). Large companies offer employees relaxation spaces varying from deck chairs to sleeping capsules. However, portable, cost efficient and moreover also effective infrastructures have not been implemented yet. Virtual Reality (VR) applications meet these requirements providing a immersive environment. Several studies support relaxation potential of VR

(Kosunen et al., 2016; Riva et al., 2007; Soyka et al., 2016). In this study we tested the effect of four relaxation conditions on mood and perceived stress.

**Hypotheses and Research Question:** H1 Independent of the method a relaxation phase increases positive affect (H1a) lowers negative affect (H1b) and perceived strain (H1c). H2 An immersive VR application enhances positive affect (2a) and lowers negative affect (2b) compared to an audio only application. H3 An immersive VR application lowers perceived strain compared to an audio only application.

**RQ1:** How does a professional immersive sleep capsule compared to a regular deck chair affect perceived relaxation effectiveness? **RQ2:** How stable are short-term relaxation effects and does this stability depend on the used relaxation method?

**Method and Data:** We conducted a 2 (VR vs. No VR) x 2 (sleeping capsule vs. deck chair) experimental within-between-subjects design. Relaxation was supported by music via headphones. Participants were assigned randomly and attended the same respective condition twice within a fixed delay of one week. After a brief information participants filled out a first set of questions, followed by a relaxation period of 15 minutes. During relaxation we obtained bio-feedback via a heart rate wristband to include an objective perspective. Afterwards, the second part of the questionnaire was assessed. We captured mood with the German version of the Positive and Negative Affect Schedule (PANAS) by Krohne and colleagues (1996). We obtained strain by using the state subscale of the German version of the State Trait Anxiety Inventory (Laux, Glanzmann, Schaffner, & Spielberger, 1981). Besides demographic data, we also assessed simulator sickness, to control for biases, and qualitative feedback. Data consists of 61 data sets virtually equally distributed to all four conditions. The average age was 22.97 (SD = 5.79). 68.9% were students and 14.8% in employment. Low scores for simulator sickness in both points of measurement were found (M1 = 1.43, SD = .44; M2 = 1.33, SD = .35).

**Results:** To examine the relationship between relaxation and positive affect (H1a), negative affect (H1b) and strain (H1c) we conducted an ANOVA with repeated measures. Contrary to the assumptions positive affect decreased after applying the relaxation ( $F(1, 57) = 7.26, p = .009, \eta^2 = .113$ ; Pre:  $M = 2.91, SE = .08$  Post:  $M = 2.74, SE = .10$ ). Additionally, there was a significant decrease of negative affect ( $F(1, 57) = 17.01, p < .001, \eta^2 = .230$ ; Pre:  $M = 1.31, SE = .05$ ; Post:  $M = 1.20, SE = .04$ ). However, an interaction effect yielded that this effect was not stable between the two points of measurement ( $F(1, 57) = 17.09; p < .001, \eta^2 = .231$ ; Pre 1:  $M = 1.38, SE = .06$ , Post 1:  $M = 1.14, SE = .03$ ; Pre 2:  $M = 1.22, SE = .06$ , Post 2:  $M = 1.26, SE = .06$ ). Strain decreased significantly after the relaxation ( $F(1, 57) = 45.370, p < .001, \eta^2 = .443$ ; Pre:  $M = 1.99, SE = .05$ , Post:  $M = 1.77; SE = .05$ ) confirming H1c. VR had no significant effect on neither mood (i.e. positive (H2a) nor negative affect (H2b)) nor perceived strain (H3).

Conclusively, we found no effect of positioning condition (sleeping capsule vs. deck chair) (RQ1). The effects were stable in between the two sessions for strain and positive affect, whereas negative affect showed deviations (see interaction effect H1b) (RQ2).

Qualitative data suggests potential benefits of customizable VR-interfaces. Objective data will be added to validate self-reported findings and provided within paper session.

**Conclusion:** Results indicate that, independent of method, relaxation balances affect leading to a more neutral state. Neither visual surrounding nor physical positioning influenced relaxation effectivity.

Future research should focus more on objective measurements and performance tests after relaxation. Additionally, new applications enabling individual customization of the content should be tested to examine the influence of personal control.

*Carolin Straßmann (University of Applied Sciences Ruhr West)*

## R-5

### Relaxing in Virtual Reality: How Does It Work?

Virtual reality (VR) has been effectively used to decrease stress and anxiety, and to increase well-being (Diemer, Muhlberger, Pauli, & Zwanzger, 2014; Kazzi et al., 2018). In fact, considering the efficacy of VR-based treatment for anxiety and related disorders (Carl et al., 2018), this approach has been considered as an alternative to an in vivo exposure for relaxation routines, especially when in vivo exposure is not feasible (e.g., the doctor cannot meet the patient), and the inclusion of relaxation routines is a mandatory part of the treatment procedure. Relaxation can be defined as a decrease in the intensity of activation of the physical and mental dimensions (Andersen et al., 2017), nevertheless, neither this definition includes any reference to any physiological parameter value, nor does it indicate how long does a patient takes to attain it. Indeed, there is a lack of studies indicating which physiological parameters are associated with relaxation perception, how much time is needed to reach this perception and/or the empirical validation of its duration (Andersen et al., 2017).

We aim at studying physiological responses that best characterize the relaxation perception induced through a VR task and the required duration for a decrease in biological activation.

36 healthy young adults (mean age  $22.3 \pm 4.1$ ; 30.6% females) who meet the inclusion criteria—24h abstinence from physical exercise, and pharmacological, alcohol and stimulants intake—and signed their consent for participating in the study, were asked to sit on a standard chair, in a soundproof room with controlled temperature (20/22°C), and to wear a VR headset displaying a 3D sitting room with a radio from where the relaxation instructions were delivered. The participants were fitted with sensors to record heart (HR) and respiration rate (RR) (Biopac MP36 system). After a 5 minute baseline evaluation—where the participants remained seated, awake, in silence, and without any stimulus—they were asked to accomplish, during one minute, a VR memory game that consisted in matching colors after flipping two pieces with a joystick - the activation period. This task was followed by a 13 minute guided relaxation task - the relaxation period. At the end of exposure, the participants completed the Self-Assessment Manikin (SAM) for pleasure, arousal and dominance, a questionnaire about the relaxation task (namely perception of relaxation) and about the VR experience, along with self-reports for presence, immersion and cybersickness. 3 participants were excluded for not meeting the inclusion criteria.

HR and RR relative to the baseline period were measured during the activation and relaxation periods. Prior to rate calculation, a high-pass filter of 1Hz was used for heart rate data and a band pass filter of .05-1Hz for respiration. The statistical analysis was based on repeated measures ANOVA. The aim was to explore whether there were differences between the baseline (first minute) and the activation period (second minute) with each minute of the relaxation task (in a total of 13 minutes), while controlling for caffeine consumption. Data from the SAM scales, presence, immersion and cybersickness were analyzed. The relaxation perception was also assessed.

A significant relationship between pleasure and dominance of the SAM scales with presence levels ( $p < .05$ ) was found, suggesting that higher levels of pleasure and dominance were associated with higher presence levels. No significant relationships were found for immersion and cybersickness. Also, results showed significant differences both for the heart rate - HR ( $p < .001$ ) and respiration rate—RR ( $p < .001$ ). Multiple comparisons with Bonferroni correction revealed that activation increased both HR (8,8%) and the RR (9,8%) relative to the baseline. After 1minute of relaxation, it was observed a decrease ( $p < .05$ ) in HR (0,7%) and RR (5,1%). The major variations occurred during the second minute of the relaxation routine (that corresponded to the VR breathing suggestions) with a decrease in 5,7% HR and 36,2% RR (all  $p$ 's  $< .05$ ). No differences to the baseline were found beyond the third minute of the relaxation task, nor caffeine consumption interacted with these results. 97.2% reported feeling relaxed, specially the initial stage of the task. The participants emphasize, as main contributions for the relaxation, the voice/instructions and the breathing suggestions, which concur with the physiological data.

VR may be effective in promoting relaxation. It may become an additional resource to the therapist's work when intervening in anxiety disorders and/or when homework is assigned to patient.

Based on peripheral physiological responses acquired through a VR task followed by a relaxation event, an estimation of the duration required to induce relaxation was found. Further studies should focus on: (A) control the effect of relaxation by using three groups: 1. - activation + no relaxation event; 2. - activation + relaxation event via sound; activation + relaxation event via avatar (reproducing the same instructions as in 2.); (B) in participants with anxiety disorders.

*Pedro Gamito (HEI-Lab/ULHT)*

## R-6

### Virtual Teams Review: An Exploration of the Multifaceted Nature of Team Virtuality

In a survey of 1,700 knowledge workers, 79 percent reported working in distributed teams (Ferrazzi, 2014). Due to the ubiquitous nature of technology in team collaboration, virtuality has been argued to be a key characteristic of all teams (Martins, Gilson, & Maynard, 2004). Many reviews have been conducted on how to define team virtuality (e.g., Foster, Abbey, Callow, Zu, & Wilbon, 2015; Hosseini, Zuo, Chileshe, & Baroudi, 2015; Orhan, 2017) but not on how to measure team virtuality. In fact, the lack of clear measurement of virtuality has been highlighted as a key methodological concern in the past decade of virtual teams research (Gilson, Maynard, Young, Vartiainen, & Hakonen, 2015). The purpose of the current study is to address this limitation by conducting a systematic qualitative review of the measures used in the virtual teams literature. The scope of the review spans from the publication of a seminal review on virtual teams in 2004 (Martins et al., 2004) to the present. By evaluating the trends in the literature over the past 15 years, we are able to summarize, critically evaluate, and provide recommendations on how to more effectively assess team virtuality.

This review was conducted using a three-step process (see Figure 1). The number of articles retrieved were reduced to a feasible size by subjecting them to the following inclusion criteria. Articles were only included in the review if they were published on or after the publication of Martins et al. (2004), were published in

a peer-reviewed journal, used empirical quantitative data collection methods, included either a manipulation or measurement of virtuality, associated virtuality with predictors or outcomes at the team level of analysis, and included team members that were human beings as opposed to automated intelligent agents. After applying these criteria to the search results, a final tally of 117 empirical articles were retained.

Based on debates in the literature on whether or not virtuality should focus on variations in communication media/technology (e.g., Kirkman & Mathieu, 2005; Shekhar, 2006), variations in distance (e.g., Chudoba, Wynn, & Lu, 2005; O'Leary & Cummings, 2007), or a combination of both (e.g., Bell & Kozlowski, 2002; Hoch & Kozlowski, 2014; Schweitzer & Duxbury, 2010), one of the attributes used to classify the virtuality measures was whether the measure focused on communication or distance dimensions of virtuality. Measures in the virtual teams literature also vary on whether they capture objective properties of the team (e.g., O'Leary & Cummings, 2002) or team members' perceived experiences (e.g., Chudoba et al., 2005). Based on this, another key attribute used to classify the measures was whether the evaluation was objective or perceptual. Even though the measures catalogued were quite diverse, this classification scheme enabled us to code the majority of articles into one or more of the following categories: objective-communication measures (64 articles), objective-distance measures (44 articles), perceptual-communication measures (35 articles), and perceptual-distance measures (5 articles) (see Table 1 for examples).

In addition to describing the wide array of measures used in the virtual teams literature, this qualitative systematic review provides a critical evaluation of each of these measurement types. Concerns with the objective-communication measures include the reliance on theories that oversimplify the experience of users (e.g., media richness theory; Daft & Lengel, 1984). What may be perceived as a more virtual medium to one team member may not be perceived the same way to another (see channel expansion theory; Carson & Zmud, 1999). Instead of viewing virtuality as an objective property of communication media we should view it as a subjective experience of its team members. This more appropriately accounts for the different experiences of digital natives and digital immigrants in the workforce (Colbert, Lee, & George, 2016) and is more consistent with the evolution of theories on team virtuality (e.g., Carson & Zmud, 1999).

Concerns with objective-distance measures include the fact that objective dispersion can contribute to perceived psychological distance (see construal theory; Wilson, Crisp, & Mortensen, 2013). Based on construal theory, there is a positive relationship between geographic dispersion and psychological distance (Wilson et al., 2013). In other words, the higher the level of objective distance in the team, the higher the perceived distance will be in the team. In order to better establish the relationship between these two type of measures, it is important to evaluate both objective and perceptual measures of distance in virtual teams.

Concerns with perceptual-communication and perceptual-distance measures include their focus on one of the key components of virtuality. Both technology and distance work together to create the virtual team experience (e.g., Bell & Kozlowski, 2002; Cohen & Gibson, 2003; Ganesh & Gupta, 2010; Griffith, Sawyer, & Neale, 2003; Hinds & Bailey, 2003; Kratzer, Leenders, & Van Engelen, 2006; Schweitzer & Duxbury, 2010). Based on these and other limitations, recommendations for the future assessment and research on virtuality are provided.

*Stefan Horgas (Pennsylvania State University)*

## R-7

### Examining the Utility of Presence:

#### 20 Years of Presence Research and Future Directions

**Statement of the Problem:** Since its rise to academic popularity in the mid-90s, presence has become one of the most widely used and studied terms in virtual reality (VR). However, conceptual and operational ambiguity around the term, as well as inconsistent usage, have resulted in an increasing obscure and fuzzy construct that is understood more heuristically than systematically.

**Analyses and Procedures:** We search the presence literature between the early 1990s and 2019 to reflect on the utility of the concept in advancing our collective understanding of VR. This literature spans work across the fields of communication, psychology, computer science, human-computer interaction, and marketing, among others. The literature is rife with varying conceptualizations of presence, ranging from "the state of non-mediation" (Ijsselstein, 2003) to "being there" (Biocca, 1997), "the psychological sense of being in the virtual environment" (Slater & Wilbur, 1997), and Lombard and Ditton (1997) present at least six "interrelated but distinct conceptualizations of presence" and settle on defining presence as the "perceptual illusion of nonmediation." These definitions overlap insofar as they refer to a state wherein users view the virtual environment as naturalistic and uninterrupted extension of the physical world. However, all of the definitions assume that users desire to experience virtual worlds as high-fidelity proxies of physical worlds, thus restricting how we think about user experiences in virtual worlds.

**Results:** Guided by discourse on the importance of constructing meaningful general theories (Shoemaker et al., 2004), we have developed three provocations to instigate discussions leading to the development of a general theory on user experiences in VR:

**Provocation 1:** Presence has become ill-defined. Presence is a multi-dimensional concept, often categorized into spatial presence (e.g., "I felt I could touch objects in the environment"), self presence (e.g., "I felt that what happened to the avatar really happened to me"), and social presence (e.g., "I felt that the interaction with the other avatar was real"). Scholars argue that the concept is platform-agnostic, not confined to virtual worlds, and earlier research has demonstrated presence perceptions from books, films, and even text messages (Lee, 2004). Presence is also argued to be both cognition and emotion (Schubert, 2009). It can be felt alone and with others (Biocca, 1997). Over decades, the presence literature has become enriched but also disordered. The lack of precise boundary conditions hinders theoretical advancement in VR because it is difficult to know what is being tested, what to expect, and whether the findings confirm or disconfirm what is known about presence (Shoemaker et al., 2004; Walther, 2009).

**Provocation 2:** Consider alternative measures of user experience in VR. One approach to conceptual clarity may be to focus on the first-hand, experiential aspect of VR content. VR is the first of its kind medium that requires users to use movements and interactions that remind them of their physical world (Biocca, 1997). Therefore, presence may simply be the result of users (a) drawing expectations of movement (e.g., I move around in a certain way) and interactions (e.g., the world responds to my movements in a certain way) from the physical world, and (b) perceiving how well the virtual world met those expectations (cf. Pettet et al., 2010). There may be alternate explications, but studying what lies at the core of user experiences in VR may help to gain greater clarity and accuracy in what scholars are pursuing when they use the term presence.

Provocation 3: Presence is not VR's end game. Another way toward conceptual precision of presence may lie in recognizing that presence may not be the ultimate goal of all VR experiences. As with other mediated experiences, the in-situ experience of presence is less important than the cognitive, behavioral, and emotional outcomes that result from the experience. Moreover, inducing feelings of presence might not always result in desired outcomes. For example, attempts at immersive storytelling increased feelings of presence in the space of a virtual world, but had no influence on users' engagement in the narrative (AUTHORS1). Engagement with the environmental features of the virtual world also hindered information processing and led to lower learning (AUTHORS2). For many types of applications, users might prefer fantasy virtual worlds that inspire awe and imagination (Possler et al., 2018) or allow for non-natural movements and interactions (Won et al., 2015) because they overcome rather than replicate limitations of corporeal space.

Conclusion: Scholars (e.g., 1992) first envisioned presence to be the defining factor of VR but its increasing popularity has led to muddled conceptualizations that restrict our understanding of virtual experiences and hinder theoretical advancement. Although it may be premature to argue for its termination, efforts to (1) add precision to its conceptual delineation, (2) consider alternative measures of studying user experiences in VR, and (3) evaluate the utility of the concept of presence may be necessary to maintain the forward momentum of the field.

*Sun Joo (Grace) Ahn (University of Georgia)*

## Wearable Devices

### S-1

#### Using Information Technology to Bridge Physical and Mental Health Indicators: Preliminary Findings on Using AI Analytics to Support Integrated Care

Utilization of information technology to promote health care, health education, and health behavior has dramatically increased in the last two decades (Lintonen, Konu, & Seedhouse, 2007; Pagliari et al., 2005), but its application in the mental health field is limited. In the health care field, new devices and smart-phone applications have been developed to assist in in vivo tracking of health indicators, such as heart rate, body temperature, exercise/activity level, and diet (Canhoto & Arp, 2017). By integrating health analytics, the application of information technology has been shown to benefit people's daily life and general health (Kostopoulos, Kyritsis, Ricard, Deriaz, & Konstantas, 2016). Wearable devices utilizing health analytics from AI platform are showing promising opportunity to offer accurate health warnings to detect physical illness (e.g., heart disease and sepsis), offering physicians more accurate diagnosis and informing clinical decision making (Glasgow, Klesges, Dzewaltowski, Estabrooks, & Vogt, 2006). In contrast, application of health analytics in the mental health field is at its nascent state, with the majority of existing applications targeting detection and diagnosis of mental health conditions (Shatte, Hutchinson, & Teague, 2019). Only recently are researchers beginning to discuss the utilization of predictive analytics for mental health (Hahn, Nierenberg, & Whitfield-Gabrieli, 2017), and many underscored the importance of including mental health researchers and clinicians to work with technology and data analytic experts (Kotz, Lord, O'Malley, Stark, & Marsch, 2018).

The emerging field of health psychology emphasizes the interconnection between physical and mental health. Research in the last two decades demonstrated the benefits of improved mental health on physical health outcome, by way of improving patient engagement, treatment compliance and treatment success of different physical conditions, such as diabetes (Bogner, Morales, de Vries, & Cappola, 2012) and heart disease (Dunbar et al., 2012). Consequently, health insurance companies and healthcare providers recognized the advantages of the integrated care model with collocated physicians and mental health providers, especially in the community setting. Despite increasing recognition of the value for integrated care, limited progress has been made in the utilization of AI technology that offers data inclusive of both mental and physical health indicators with the goal of promoting holistic well-being.

Given the close relationship between heart disease and anxiety symptoms, research utilizing AI technology to support holistic well-being, integrating mental and physical health indicators, may start with a measure of "mental distress" that consists of both physiological and psychological indicators. Using an autonomic nerve detector, the traditional measure of stress captures the degree of continuous heart rate change through heart rate variability (HRV) analysis to assess the balance of the autonomic nervous system (Jovanov et al., 2003). At the most basic level, HRV is an indicator of overall health that captures both mental distress and physiological indicators of heart diseases. Incorporating an existing health analytics AI platform, the current study aims to examine various neurophysiological indicators, including BMI, heart rate, and sleep as input into an analytic algorithm to offer evidence-base information reflective of the mental distress index (i.e., HRV) that can be tracked moment by moment. Data were collected to offer preliminary analysis on the utility and future development of the index.

Methods and Results: Using the NeuroSky instrument with a health analytic algorithm (CURA, Inc.) that captures heart-rate data continuously to calculate HRV, 260 healthy individuals entered into the study between 2016-2017. Participants include 83 males (32%) and 177 females (68%) with a mean age of 40.4 (SD = 8.3). Individuals with a history of heart disease and mental health diagnoses were excluded. Preliminary analysis explored factors related to mental distress, including BMI, heart rate, and sleep. Results indicate that, comparing to those with lower BMI ( $\leq 18.5$ ), individuals with higher BMI ( $\geq 25.0$ ) tends to have higher level of mental distress ( $T = 5.75, p < .001$ ) and higher heart rate ( $T = 11.76, p < .001$ ). Additionally, individuals with higher levels of mental distress likely sleeps less than those with lower stress level ( $T = -3.7591, p < .001$ ).

Conclusion/Implication: Using the AI technology to bridge physical and mental health indicators can support ongoing integration of mental and physical health care. The current study suggests that HRV as an indicator of mental distress shows some promising feature in its utility to capture signs of mental and physical health, such as BMI, sleep, and heart rate. Future application that aims to develop technological support integrating health and mental health indicators may consider designing appropriate support network in consultation with clinician experts. The health analytic technology can be integrated into a smartphone application to coach people on managing their overall well-being via life style change. For instance, the technology may offer text message reminder for users to get some sleep or video call for users to receive social support as needed depending on individualized profile with input from in vivo life data.

*Mengchun Chiang (Carnegie Mellon University)*

## Workplace / Future of Work

### T-1

#### **Welcome to the (Email) Machine: Exploring the Relationship Between Electronic Communication Delivery and Work/Home Boundaries**

With the rise of artificial intelligence and machine learning, experts (e.g., Chamorro-Premuzic, 2016; Kruse, 2018) have suggested that robot managers are closer to widespread use than previously thought. This will give businesses more flexibility with respect to management and could drastically alter how organizations interact with their employees. Yet this potential has largely been suggested. At this point in the development of automated management approaches, we have little evidence of how these tools impact the ways in which companies communicate with their employees nor how employees might respond to such communication.

As it has been argued, the rise of mobile and sophisticated information-technology infrastructures within organization has brought with it the potential to distort employees' perceptions of work-life balance. That is, work is now increasingly creeping into spaces that have been off-limits from work (Gregg, 2011). Affective events theory (AET; Weiss & Cropanzano, 1996) is one theory that can help explain this. AET suggests that work-related events lead to emotional responses and subsequent impact on job outcomes such as job satisfaction and organizational commitment (Mignonac & Herrbach, 2004; Scott & Judge, 2006). One such event that has also been studied is the timing of work e-mail communication (Walther & Tidewell, 1995) and how it impacts work-nonwork conflict (Butts, Becker, & Boswell, 2015). Overall, a large body of scholarly research has considered the ways in which individuals navigate the boundaries between work and home (see Berkelaar & Tronstad, 2017; Kirby & Buzzanell 2014 for recent reviews). Often, research on work-life balance has examined how individuals communicate their expectations for boundaries between work and home, often focusing on the relationship between communication technology use and work-home conflict (Wright et al., 2014). As the rise of automated management systems and artificial intelligence-based bosses may exacerbate these issues by shifting both the timing of communication with employees and perceptions related to work-life balance. The current study will examine some perceptions based upon this work-life timing, but also related to the nature of the source of the message.

**Method:** An online experiment is being pilot tested (data collection and analysis will be completed by the conference) to address the hypotheses and research questions. The study employs a 2 (time) x 2 (source) between-subjects design in which participants read and respond to two e-mails from a manager to an employee; first, an e-mail making a request of the employee, and second, a follow-up e-mail making the request again. The second e-mail will have a time stamp suggesting it was sent either during work hours or during non-work hours, and it will also have a signature line saying it was automated or not.

After reading one of these four pairs of e-mails, participants (recruited through on-campus convenience samples and social media) will then be asked to respond to several measures examining how they think the employee would perceive the message and what they think the nature of the relationship between the manager and the employee. Participants will first respond to measures of the perceived job satisfaction (Eisenberger, Cummings, Armeli, & Lynch, 1997) and affective organizational commitment (Allen & Meyer,

1990) of the employee. Drawing from Walther and Tidwell (1995), participants will also be asked about the perceived intimacy and dominance in the manager-employee relationship, using Burgoon and Hale's (1987) relational communication questionnaire. Finally, participants will be asked to respond to how they think the employee will feel when reading the second e-mail using the PANAS (Watson, Clark, & Tellegen, 1988) and standard demographic questions that also include question asking about their own employment. Analysis will focus on using ANOVA to test for differences in outcome measures based upon condition. Data collection and analysis will be completed before the conference.

*David Westerman (North Dakota State University)*

### T-2

#### **Understanding the World of Work: Machine Learning Approaches to Clustering the O\*NET Database**

Understanding the future of work requires understanding the nature of it. To that end, the Occupational Information Network (O\*NET) is an extensive occupational database, developed and updated by the US Department of Labor, covering virtually the entire world of work. Specifically, the O\*NET contains 968 classes of occupations that are described by 277 characteristics reflecting the nature of work (e.g., occupational requirements, workforce characteristics, occupation-specific information) and the nature of the worker (worker characteristics, experiences, and requirements).

Although there have been a few past attempts at empirically summarizing this database, we know of no empirical attempts of applying machine-learning (ML) algorithms that might extract additional useful information about the nature of occupational clusters, and what occupations and their clusters might predict (e.g., educational requirements, job prospects, average salary, degree of automation).

The poster as proposed will present several ML approaches toward occupational clustering of the O\*NET data (e.g., k-means, hierarchical clustering, Gaussian mixture modeling) and prediction of outcomes with these data (e.g., random forests, support vector machines, elastic net). Results will be compared against traditional methods of clustering (e.g., Ward's method, k-means without cross-validation) to demonstrate the potential utility of ML. The work is also intended to inspire future ML applications to occupational data that serve to improve our understanding the current world of work, so that we can better contribute to the future of it.

*Fred Oswald (Rice University)*

### T-3

#### **Industry 4.0.: Mapping Emergent Job Design and Teamwork Dimensions in Units Working with Artificial Intelligence Technologies**

Businesses are currently entering the world's fourth industrial revolution, aptly named Industry 4.0. Characterized by new levels of connectivity between workers and machines, access to unprecedented amounts of data and the use of innovative technology to accomplish work-related tasks, Industry 4.0 has crucial financial and societal impacts (Schwab, 2017). In Canada, 1.1 million jobs are created, and 4.8 billion dollars are invested in research and development each year by Industry 4.0 companies (ITAC, 2015).

At the forefront of Industry 4.0 is the increasing implementation of artificial intelligence (AI) enabled technologies among workplaces. For example, about sixty percent of trades on Wall Street are now solely executed by AI (Accenture Applied Intelligence, 2018). As much as some fear the replacement of human workers by computationally advanced machines, others see AI as an opportunity for workers to benefit from more effective and enriching work settings (Brynjolfsson & Mitchell, 2017; Frey & Osborne, 2013). Anchored in this second perspective, our first research question is: what are the new dimensions of work, in terms of job design and teamwork, brought by the fourth industrial revolution and the use of AI? Our second question is: are there validated instruments available to measure these emerging dimensions?

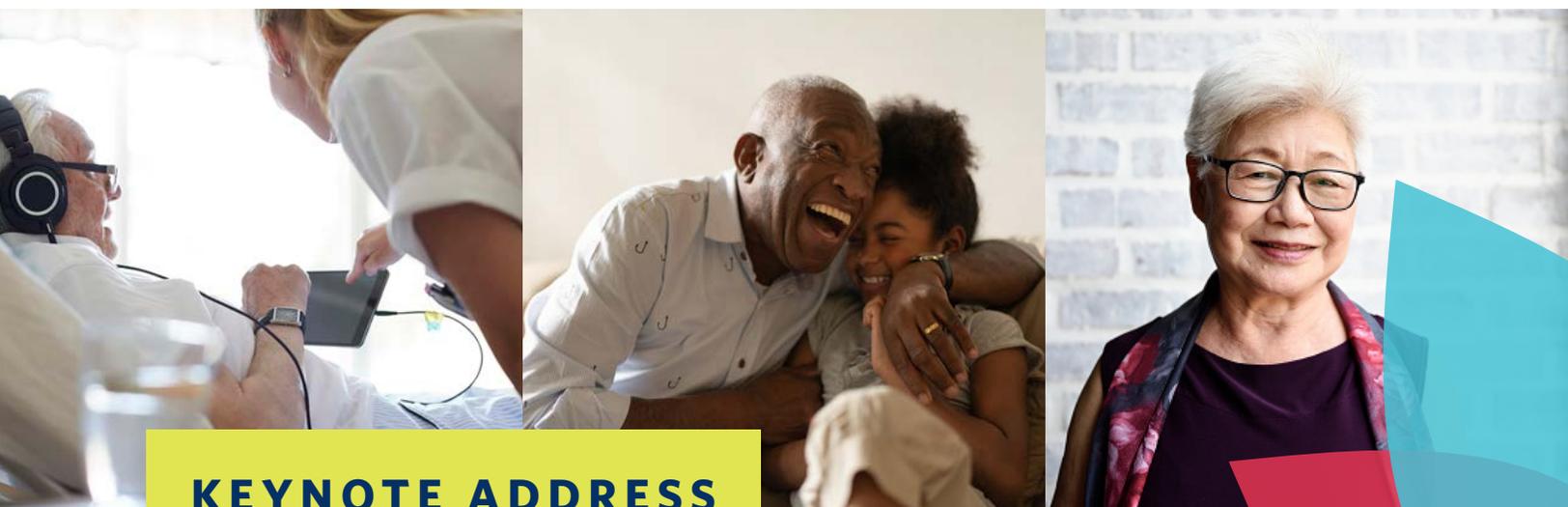
To answer our research questions, we conducted a literature review (until October 2018) using a two-step approach. First, we searched the nascent literature on Industry 4.0. to find the main articles identifying emerging dimensions of job design and teamwork processes (e.g., Battistelli, & Odoardi, 2018; Cascio & Montealegre, 2016; Parker et al., 2017; Schwab, 2017; Swartzmüller et al., 2018). Our findings were then classified as pertaining either to job design or teamwork processes (other types were not considered). A dimension was classified within the job design category when it involved the configuration of tasks specific to a position and its perceived outcomes on workers (Oldham & Fried, 2016; Parker et al., 2017). On the other hand, a dimension was classified within the teamwork category when it referred to dynamic interactions between two or more entities (e.g., human-human, human-computer) evolving in a working system (Kozlowski & Chao, 2018). In parallel, we also performed a thorough search on the PsycINFO database with the keywords “artificial intelligence” or “fourth industrial revolution” and “work\*” to identify additional works. Our search resulted in the identification of 57 relevant documents throughout which we first looked for emerging dimensions of work. To be considered emergent, a dimension had to be distinct from those already listed in mainstream job design (Morgeson & Humphrey, 2006) and team composition/processes models (Bell et al., 2018; Grossman et al., 2017). We then reviewed the 57 documents to determine if a measurement instrument had been created for each of the dimensions identified.

We identified eleven emerging dimensions pertaining to job design and teamwork brought by the fourth industrial revolution and IA. Six of them were classified as new job design dimensions: 1) the strategic character of tasks (not easily replaceable by IA), 2) the augmented nature of human-computer collaboration, 3) the quality of person-machine interactive interfaces, 4) work accomplishment mobility (time, physical space), 5) performance data monitoring, and 6) the ethical utilization of data. Five emerging issues concerning teamwork dynamics were found: 1) trust in the learning machines, 2) the exchange of information between humans and learning machines, 3) the attribution of responsibility between learning machines and workers, 4) justice in the appreciation of collaborative human-machine performance, and 5) collaboration with ad hoc entities and partners in the digital space. As for measurement instruments, we found validated scales to measure only two out of the six emerging job design dimensions (Alge, 2001; Brougham & Haar, 2018) and only one out of the five emerging team dimensions (Charalambous et al., 2016). Finally, we noted that there were very few empirical studies on the emerging issues identified in the Industry 4.0. literature.

In light of this literature review, two conclusions can be drawn. First, the support for new job characteristics and team dynamics

pertaining to the arrival of AI in the workplace remains largely theoretical. Furthermore, there are no instruments to measure most emerging dimensions identified in our review. Most issues in the Industry 4.0. have been anticipated or proposed from a conceptual standpoint rather than from empirical studies. Based on this review, we recommend the adoption of a mixed method approach to researchers interested in mapping emerging job design and team dynamics. Qualitative work is needed to corroborate that emerging issues identified in this review are perceived as relevant by individuals working with AI, and that no other issues have been missed. Then, a quantitative phase will allow the development of reliable and valid scales to measure the emerging dimensions and assess their utility in comparison to current job and team design models.

*Viviane Masciotra (University of Montreal)*



## KEYNOTE ADDRESS

9:15–10:30 A.M.

**Independence Ballroom**

# Aging and Technology Systems: Benefits and Challenges

This keynote address will explore how technology can help people successfully age, including staying healthier, more active, and socially connected, and being able to continue to contribute to the workforce into older age.

Sara J. Czaja, Professor of Gerontology and the Director of the Center on Aging and Behavioral Research at Weill Cornell Medicine, will discuss how technological systems can be used to support the independence and quality of life of older people and their families, how these systems can be designed to be commensurate with the needs, characteristics, and preferences of these user groups, myths and realities of older people and technology, and recommendations for various stakeholders including caregivers, health care professionals, designers and developers, employers, and policy makers.



## **SARA CZAJA, PhD**

### **Director**

*Center on Aging and Behavioral Research in the  
Division of Geriatrics and Palliative Medicine  
Weill Cornell Medicine*

### **Professor of Gerontology**

*Weill Cornell Medicine*

10:45 A.M.—12:00 P.M.

## Concurrent Sessions 1

### Lafayette Park

## Games and Virtual Reality For Fostering Education and Knowledge

### PAPER SESSION

#### **An Evaluation of TMM ARventure: An Augmented Reality Game for a Science Museum in the Philippines**

Augmented Reality (AR) refers to the superimposition of interactive digital information on real settings in 3D space to provide users with supplementary situated knowledge (Azuma, 1997). Within educational contexts, AR applications are created and deployed to serve affective and cognitive outcomes. They have been shown to increase student motivation (Dunleavy, Dede, & Mitchell, 2009), enjoyment, engagement, and interest (Akçayır & Akçayır, 2017), and to foster conceptual understanding in science, mathematics, engineering, and technology (STEM) fields in particular (Ibanez & Delgado-Kloos, 2018; Woods, Billingham, et al. 2004).

These positive outcomes, however, come with caveats. AR applications can lead to cognitive overload (Akçayır & Akçayır, 2017), may increase distraction (Ibanez et al., 2018), and foster unintended behaviors such as a misplaced sense of competition that shifts focus from learning to “beating” other teams (Dunleavy, Dede, & Mitchell, 2009). Furthermore, AR applications are difficult to create (Woods, et al., 2004) and are logistically challenging for teachers to deploy (Akçayır & Akçayır, 2017).

Are AR applications worth the trouble? Do their benefits outweigh their costs? The purpose of this paper is to evaluate an AR game designed for an informal STEM learning environment, to determine its impact on the cognition and affect of its users.

The Mind Museum (TMM) is a first-class science museum located in Manila, Philippines. Its mission is to provide the public with extraordinary educational experiences that lead to a deeper understanding of science. The museum has over 250 interactive exhibits in five galleries.

In 2017, TMM’s parent organization, the Bonifacio Art Foundation, Inc., entered into an agreement with the Ateneo de Manila University to create an augmented reality game to increase visitors’ engagement by providing them with a sense of mission and adventure as they explore the exhibits.

The result of this collaboration was TMM ARventure. This story-based adventure game is divided into three main chapters. In Earth Adventure, the player assists an alien visitor in repairing her crashed spaceship and studying Earth by looking at exhibits about the planet. At the start of the second chapter, Tiny Adventure, a piece of equipment has shattered into fragments. To recover the pieces, the player has to visit exhibits about atoms and microorganisms. Finally, in Future Adventure, the player assists a time-traveler recover data from a broken cellular phone by going to technology-themed exhibits.

As of the time of this writing, TMM ARventure is undergoing field testing. We are planning a field study in which we will ask senior

high school students to play the game as we collect data along two dimensions: engagement and learning. To measure engagement, we will be examining the usage logs, specifically the time to complete each chapter. We will ask players to complete two questionnaires: the Handheld AR Usability Test (HARUS; Santos et al., 2015) which measures the comprehensibility and manipulability of the AR experience and the game experience questionnaire (Jsselsteijn et al., 2013) that measures the extent to which the users enjoyed the game. To measure learning, we will ask players and non-players to complete a short comprehension test to determine if they noticed certain details from the exhibits. The results from the two groups will be compared.

From this evaluation, we hope to estimate TMM ARventure’s value as an informal learning tool and as an enrichment to the museum experience. More broadly, we hope to contribute to what is known about the design of AR applications for informal settings by identifying the factors that helped or hindered the application from achieving its desired outcomes.

*Ma. Mercedes Rodrigo (Ateneo de Manila University)*

#### **Improving Student Attitudes and Performance in STEM through Virtual Reality and Constructive Feedback**

More jobs than ever require proficiency in science, technology, engineering, and math (STEM; Rothwell, 2014). Consequently, understanding how to bolster engagement in and attitudes toward STEM is a critical initiative for researchers and educators. The development of immersive educational technologies such as virtual reality (VR) present a potential avenue to achieve greater STEM engagement. VR allows students to learn and practice STEM skills in an immersive environment, which research suggests is an effective medium for the transfer of STEM skills and knowledge (Parong & Mayer, 2018). What remains largely unexamined is the degree to which VR can increase student engagement in and improve student attitudes toward STEM careers.

We address this research need by empirically examining two questions. First, we asked how using VR might influence student attitudes toward STEM as compared to traditional pedagogical methods. Second, we investigated how the presence of others affects attitudes and performance through the provision of peer feedback. While countless educational studies exist demonstrating the general importance of collaborative learning (Johnson & Johnson, 2009), research has yet to examine this phenomenon in the context of STEM education in virtual environments. We conducted two studies to investigate these questions in the context of welding education. Welding incorporates STEM content including science, technology, and math (Asunda, 2012). VR welding simulators are increasingly used in welding education and allow for less resource-intensive and safer skill acquisition (Byrd et al., 2015).

To address our first research question, we conducted an experimental study in which undergraduate students (N= 122) were randomly assigned to one of five welding education task conditions: literature, video, low-fidelity simulator, short-exposure VR (5 minutes), and long-exposure VR (10 minutes). These conditions ranged from least immersive technology (literature) to most immersive

technology (VR). To begin, each student watched a brief contextual video about welding. Participants then completed their assigned welding education task. Lastly, each participant completed a survey measuring attitudes toward welding including career interest, person-occupation fit, and self-efficacy.

ANOVA results suggest that VR may increase welding self-efficacy, career interest, and person-occupation fit ( $F= 2.825$ ,  $p= .028$ ). For example, long-exposure VR led to greater self-efficacy compared to literature. Qualitative survey data reported by participants who engaged with the VR welder further supports this finding. When asked whether they could become a good welder, one participant responded, “based on my performance today at the simulator, ... I could be a good welder.” Another expressed that welding is “a more involved and technically complex process than I had originally thought... it seems pretty fun.” We also examined the potential of VR to improve self-efficacy among underrepresented populations in STEM (i.e., women). Gender analyses showed that men had higher efficacy across conditions ( $t= 1.92$ ,  $p= .05$ ); however, trends indicate that VR may diminish this disparity. Thus, VR may show the potential to reduce STEM gender gaps.

To address our second research question, we conducted a second experimental study in which undergraduate students were randomly assigned to one of three social learning conditions that varied in group size (2, 3 or 4 participants). Groups began by watching a short introductory welding video. Following this video, each student had a turn to practice VR welding for 5 minutes in turn. While their peer was welding, other students were permitted, but not prompted, to provide feedback. Additionally, the experimenter was allowed to provide constructive feedback if it was asked for by participants. The experimenter tallied the frequency of participant feedback-seeking, the amount of peer feedback received, and rated the quality of feedback received for each participant. Welding performance data was also captured for each participant. Lastly, participants were instructed to complete a survey measuring personality, goal orientation, and attitudes toward welding, including career interest, person-occupation fit, and self-efficacy. Data will be analyzed to answer our second research question. Preliminary observations indicate that participants who struggle with VR welding may improve their performance and demonstrate better attitudes after seeking and/or receiving constructive feedback. Importantly, extant research in group participation leads us to believe that group size may affect the feedback peers provide (Thomas & Fink, 1963). Data collection and analysis for study 2 is ongoing and will be completed in April 2019.

Results from our studies have several implications. Social Cognitive Career Theory (SCCT; Lent, Brown, & Hackett, 1994) suggests the importance of self-efficacy in developing and pursuing career goals and intentions. Study 1 suggests that VR, as compared to more traditional teaching tools, can strengthen self-efficacy and decrease the gender gap related to STEM. Thus, introducing VR into training programs may help diversify and grow the number of students pursuing STEM. Further, preliminary results from study 2 suggest that VR training may be most effective when implemented in a collaborative environment that facilitates constructive feedback. We anticipate that final data analysis will further support this conclusion.

*Bradley Pitcher (George Washington University)*

### **Gamifying Crowdsourced Contributions:**

#### **The Impact of The Scoring Mechanism Design on Players' Behavior in a Game Designed for Pooling Knowledge- Encouraging Quantity and Quality Contributions**

Introduction: Games and game technologies increasingly transcend the traditional boundaries of their medium, as evidenced by the proliferation of serious and pervasive games (Ritterfeld, Cody & Vorderer, 2009). The most recent phenomenon in this field is ‘gamification’, an umbrella term for the use of game elements in non-game services and applications to improve user experience and engagement (Deterding et al., 2011). To date, gamification has become a popular strategic instrument to engage people in marketing as well as in non-business contexts such as politics, health, education and more (Richter, Raban & Rafaei, 2015).

One of the most popular application areas of gamification is crowdsourcing (Seaborn, Fels, 2015), where information is pooled out from a large number of users for purposes such as solving problems (Brabham, 2008), creating knowledge (Palacios et al., 2016), and processing data (Haythornthwaite, 2009). To date gamification serves as a major instruments for driving users’ motivation in crowdsourcing systems that address real world challenges such as DNA research, object recognition, building archives, tracking climate changes and more (Richter, Raban & and Rafaei, 2018). The most common purpose behind using a gamified approaches is to encourage behavior change: increase participation, improve performance, or quality of contribution. Thus, the success of these systems largely depends upon the motivation for participation of each contributor and the quality of input (Ye et al., 2012, Cooper et al., 2010). Since games are known as an effective tool in raising motivation it seems natural to integrate game elements as a motivator to contribute to knowledge pooling. Although, there are some empirical studies indicating gamification as an effective approach to increase users’ motivation and participation (Guy et al., 2011, Richter, Raban & and Rafaei, 2018, Richter, Raban & Rafaei, 2018), research gaps still exist, preventing a full understanding of the gamification potential and therefore from designing effective gamified knowledge contribution solutions.

The current research aims to: a) shed light on the effect of a particular gamification feature- the scoring mechanism design on users performance, b) investigated whether differences exist between various crowd’s types in regard to different scoring designs.

Experimental Design, Research Tool & Analyses: We implemented a control experiment to discuss how different scoring mechanism designs have a differentiated effect on user performance in terms of intensity of participation, play time utilization and quantity and quality of contribution.

The study employed a crowdsourced knowledge game, called GUESS, which was developed by IBM (Figure 1). The game prompts a questions at the rate of one question per minute. Players are asked to enter as many relevant responses to the specific question within a time frame of one minute. We compare usage of a monotonic and predictable linear function, with an exponential function which is more likely to be surprising and introduces irregularity to the system. We examined the server logs, which documented the details of each response provided by users along with a time-stamp.

Three different groups of users were examined exposing each to the different scoring conditions. One group included heterogeneous crowds - general public, included teenagers and older people. While the two other groups were more homogeneous crowds (students). We identified two student’s groups differed in the relevance of the

activity to their learning domain. One group were management students answering questions on management topics (noted as 'experts'). The other group were students from diverse discipline. We gave the same general questions that we gave the general public.

To compare results from manipulation conditions with those from control condition an HLM model (Hierarchical Linear Modeling) was conducted. A backward elimination procedure was used, whereby non-significant terms were dropped one by one and the model re-assessed to determine the significance of each of the remaining variables at every stage. We will present analysis and the results in details in the conference presentation.

**Discussion and Conclusions:** According to our findings, exponential scoring was associated with contribution of some span, however, it is also associated with gains in performance, such as quantity of contribution, play-time and participation. Given general questions, students show better performance than the general public. Yet; the hardest challenge is to pool quality responses from the general students. That said, both experts and general students are indifferent to the scoring condition. A crowd of experts is more difficult to mobilize than general crowds, yet experts are at the top of the hierarchy of crowd type for quality. Moreover, experts generate high quality responses regardless of the point allocation mechanism; incentives in the form of points do not lead to quality impairment.

This work main novelty lies in bringing together crowd-sourcing and game capabilities for different scoring mechanism designs. Our findings suggest that a trade-off does not necessarily exist between the quality and quantity of contributions, and that it could be more effective to leverage knowledge contribution behaviors by using tailored incentives.

*Ganit Richter (University of Haifa)*

---

## Farragut Square

### Bias and Technology

#### PAPER SESSION

#### DiVirtuality—Designing and Testing a Virtual Gallery for Stereotype Reduction and Diversity Awareness

Stereotypes, prejudice and discrimination are natural and prevalent phenomena in interpersonal interaction that have never lost relevance. Their mechanisms and contents raise conflicts and deeply impact individuals' and societies' good life. At the same time, diversity becomes increasingly important to organizations, due to globalization, workforce diversity and job complexity in times of digitalization. Diverse teams have shown to be better in solving complex tasks and team diversity is positively correlated to well-being (Wegge, Roth, Neubach, Schmidt & Kanfer, 2008). Therefore, it is important to raise diversity awareness (Dehaze, 2018; Jehn, Greer, & Rupert, 2008, Page, 2008). This is promoted by a change of the individuals' stereotype beliefs, perceived groups variability and positive contacts between groups. Changing perspectives, reflecting upon similarities and differences and social roles can help developing sympathy for people of different groups (Eimler, 2014). Scalable interventions that help users to do so, are needed. Therefore, in the present work

a virtual environment for diversity sensitization in form of a virtual gallery was constructed and evaluated. VR technology has shown to be effective in persuasion settings, such as environmental protection training (Ahn, Bailenson, & Park, 2014). People can get fully immersed without any distortion to their concentration. They can explore multimedia material freely, without being forced, which should reduce reactance (Brehm, 1966) and might inspire their intrinsic motivation to explore the material. Consequently, the following hypotheses are assumed: H1: Users state more positive diversity attitudes after using the diversity sensibilisation than before; H2: Users state more positive diversity attitudes, when the diversity sensibilisation was presented in a virtual reality environment compared to the tablet usage; H3: Users enjoy the use more (a), perceive higher feelings of flow (b) and are more positive affected (c) by the use of the diversity sensibilisation in the virtual reality environment compared to the tablet usage.

To test hypotheses we used a two factorial between-subjects design. Participants were assigned randomly to one of the experimental conditions (Virtual Reality vs. tablet). One week prior to the main experiment a pre-questionnaire assessed demographic information and baseline diversity attitudes. The actual diversity sensitivity training consisted of a 20 minutes exploration of a virtual gallery, either in VR or on a tablet. During this phase participants inspected the system freely. The VR application was developed with the Unity 3D engine and displayed through the Oculus Rift CV1. Participants could interact with elements focussing on behavioral (e.g. hot wire game with a tremor-bias), cognitive (e.g. facts and figures on virtual displays) or affective (e.g. video or audio sequences featuring discriminative experiences) channels.

Diversity attitudes were captured as three separate constructs: a) Universality - measured with a 9-item subscale of the Spiritual Transcendence Scale (Piedmont, 1999), b) personal beliefs, using a 15-item subscale of the Diversity Beliefs in Personal and Professional Contexts (Pohan & Aguilar, 2001) and c) motivation to control prejudices (Banse & Gawronski, 2001). Enjoyment was measured using AttrakDiff subscales for hedonic and pragmatic quality by Hassenzahl and colleagues (2003). We captured flow with the short scale by Rheinberg, Vollmeyer, and Engeser (2003). Participants filled out the German version of the Positive and Negative Affect Schedule (PANAS) by Krohne and colleagues (1996) to assess mood. All constructs showed satisfactory to good internal consistencies.

Data consists of 50 data sets (8 female and 42 male) equally distributed to the conditions. The average age is 22.1 years (SD = 4,04). 94% report on regular usage of a tablet, 52% have experienced VR before.

Results of the paired-sample t-test show a significant effect of the diversity sensibilisation leading to higher universal orientation (H1a) ( $t(49)=-5,47, p<.001$ ) and higher personal beliefs about diversity (H1b) ( $t(49)=-3,22, p=.002$ ), independent of the condition. However, there was no significant difference between pre and post measures of motivation to control prejudice (H1c). ANOVA with repeated measurement did not confirm H2 which expected a superior effect of immersive VR regarding diversity attitudes. However, results of independent t-tests show that the virtual environment appears to be more enjoyable (H3a) ( $t(48)= 2,52, p=.015$ ), triggers flow marginally (H3b) ( $t(48)= 1,940, p=.058$ ) and leads to higher positive affect ( $t(48)= 2,52, p=.015$ ) than the tablet condition.

The results indicate that the diversity sensibilisation was effective, since users showed more positive personal beliefs after the

experiment than before. Contrary to the assumption, the virtual environment did not lead to higher diversity awareness than the tablet condition. However, people enjoyed the virtual environment more, perceived higher feelings of flow and were positively affected by its usage. Thus, with regard to the positive computing approach (Calvo & Peters, 2014) the virtual gallery can be evaluate as more effective: Technologies that are more fun to use and affect the users positively might also lead to higher usage intention and, as a consequence, higher effectiveness. This link between positive affect, usage intention and effectiveness needs to be investigate in future research.

*Carolin Straßmann (Hochschule Ruhr West University of Applied Sciences)*

### Racism and Sexism by Computers Reduce Outrage

In 2014, Amazon created a machine learning bot that screened the résumés of job applicants. However, the program was scrapped in 2017 after it has been discovered that it was systematically discriminating against women. Although Amazon scrapped the program, was the backlash mitigated by the fact that a compute algorithm—not a human—was the one making the sexist hiring decisions? People certainly show moral outrage when humans discriminate. Here, we test the possibility that people show less moral outrage when computer algorithms discriminate.

There are two reasons why people might be outraged by discrimination. The first is the negative outcome—discrimination harms the population discriminated against. The second is the motivation behind the discrimination. People might be outraged not only because of the bad outcome, but also because of the discriminatory motivation behind it. Since people find it harder to attribute negative motivation to computer programs (e.g., “that algorithm hates women”), people might be less outraged at and more accepting of discrimination committed by algorithms rather than by humans. We therefore hypothesize that (1) people will be less outraged by discrimination by algorithms than by humans, and that (2) this lower outrage will be explained by the reduced negative motivation attributed to algorithms. We tested these hypotheses in 3 studies, examining whether participants react less poorly to discrimination committed by an algorithm.

In Study 1 (N = 120), we tested our basic hypothesis that people will be less outraged and more morally accepting of discrimination by machine than by human. Participants read about either a human or an algorithm that discriminated against black people in hiring decisions for engineering jobs. Participants than answered 5 items measuring their moral outrage, on a 1 to 7 Likert scale. We found that participants were less outraged when the discrimination was by an algorithm (M = 5.34) than by a human (M = 6.22,  $p < .001$ , Cohen's  $d = 0.80$ ).

In Study 2 (N = 240), we replicated Study 1 by measuring participants' responses to hiring discrimination against women. As in Study 1, participants were less outraged by a discriminating algorithm (M = 5.54) than a discriminating human (M = 6.00,  $p < .001$ , Cohen's  $d = 0.45$ ).

In Study 3 (N = 240), we tested prejudiced motivations a potential mediator of the effects observed in Studies 1 and 2. The first part of Study 3 was identical to that of Study 1. After rating moral judgment, participants answered 4 items measuring the negative motivation attributed to the agent (e.g., “... is a racist”) on a 1 to 7 Likert scale. As in Study 1, participants were less outraged when the

discrimination was committed by an algorithm (M = 5.68) than by a human (M = 6.07,  $p = .012$ , Cohen's  $d = 0.33$ ). Moreover, this effect was mediated by people's reduced negative motivation to the algorithm, such that in the model that included both the direct effect and the indirect effect, while the indirect effect was significant ( $b = .25$ , SE = .05, CI.95[.17, .34]), the direct effect was not ( $b = -.05$ , SE = .08, CI.95[-.20, .1]).

This research emphasizes the role of motivation attribution in outrage towards discrimination. We show that people are outraged by discrimination not only because of the harmful consequences, but also because of the motivation underlying the discrimination. Discrimination that is motivated by racist or sexist motivation is seen as worse. Moreover, our research raises a red flag over the speed in which some automatization processes are happening. Automatization has a huge potential. Specifically for hiring decisions, algorithms can be much more impartial than humans, leading to greater equality. However, caution is needed, as algorithms are not always as impartial as they were planned to be. People's natural defenses against injustice and discrimination are lowered when the agent is an algorithm, making it easier for discrimination to go unnoticed, unobserved and unopposed.

*Yochanan Bigman (The University of North Carolina at Chapel Hill)*

### The Development of a Virtual Reality Shooter Paradigm and Its Application to a Study on Racial Bias in Police Officers' Use of Lethal Force

Statement of the Problem: Recent media attention to police violence towards unarmed Black men and women is unprecedented. A growing body of empirical work confirms racial bias against Black compared to White suspects (e.g., faster reaction time shooting Black suspects; Correll et al., 2002; Plant & Peruche, 2005). However, critiques of this work emphasize the lack of ecological validity (James, Klinger, & Vila, 2014); participants typically sit in front of a computer and respond to still images of “suspects” who appear on the screen holding a gun or a benign object (e.g., Correll et al., 2002). These methods are not likely to induce the type of cognitive and affective states experienced by people operating in high-risk environments. In our study, we use Virtual Reality to embed participants in 360 immersive scenarios which unfold unpredictably; for example, suspects may appear from the periphery or even from behind the participant. Integrated eye-tracking allows us to examine how the officers' affective state (i.e., measured through pupil dilation) and cognitive processing (i.e., measured through attentional fixations and saccades; Nisiforou & Laghos, 2016) interact in the culmination of decisions to use lethal force. Furthermore, psychological states may differ when officers are faced with Black or White suspects. Research has found Whites' to have greater levels of physiological arousal when interacting with Black compared to White dyad partners (West et al., 2017) and the link between physiological reactivity an aggression is well established (e.g., Scarpa & Raine, 1997). Moreover, when cognitive demand is high, people tend to rely on racial cues heuristically to inform decisions (Jones & Kaplan, 2003). Thus, in this study we predict that both high physiological arousal and insufficient attention will mediate racial bias in shooting reaction times towards Black suspects.

Procedures: Scenarios were scripted and filmed with a 360 video camera and included scenes such as traffic stops and altercations in the street. Suspects and bystanders were played by volunteer theater



majors and free-lance professional actors. Each scene was filmed once featuring a White suspect and once featuring a Black suspect. We constructed two “sets” of 12 unique scenarios to which participants are randomly assigned; within each set, 6 include a Black suspect and 6 include a White suspect. Each scenario is 30 seconds long and unfolds such that a decisive stimulus (a gun or benign object) appears at the 28 second mark. Consenting participants are fitted with an HTC Vive VR headset with integrated eye-tracking capability (Tobii Technology), and given a hand controller—which is depicted in the VR as a police issued handgun—for recording decisions to shoot. Following the VR procedure, participants also complete a survey battery which includes, for example, questions about their experience in the VR. In total, the procedure takes about 45 minutes to complete, and participants receive \$10 for completing the study.

**Planned Analyses and Preliminary Results:** Currently 8 participants have completed the study and data collection is expected to be completed by the summer of 2019. Out outcome variable of interest is reaction time which is measured by a script embedded in the VR program, which records the time in milliseconds between the emergence of the stimuli (either the gun or benign object) and the pulling of the trigger. Predictor variables of interest include: physiological arousal which is measured through pupil dilation and is operationalized as the maximum pupil size achieved during each scenario (Kinner et al., 2017); eye-tracking patterns, which will be measured as the ratio of dwell time on the suspect (McDonnell et al., 2014); and suspect race which is a within-subjects variable as each participant is exposed to 6 White and 6 Black suspects. Thus far participants’ overall average reaction time was 1161.11 milliseconds, and ranged from 913.20 to 1446.94 across participants ( $SD = 182.25$ ). Participants reaction times towards Black suspects ( $M = 1143.21$ ;  $SD = 447.55$ ) and White suspects ( $M = 1179.01$ ;  $SD = 218.73$ ) are thus far not significantly different [ $F = .028$ ;  $p > .05$ ]. Eye-tracking and physiological arousal data are currently being prepared for analyses. Hypotheses will be tested using MEMORE, an SPSS macro for within-subjects mediation (Montoya & Hayes, 2017).

**Implications and Conclusions:** We expect findings to have both content-based and methodological implications. If our hypotheses are supported, we believe this will be first finding of its type to establish physiological arousal and disruptions in cognitive processing as explanatory variables for police officers’ faster reaction times when faced with Black relative to White suspects (e.g., Correll et al., 2002; Plant & Peruche, 2005). We believe VR is a promising mechanism for the study of social-psychological behavior in high-risk environments, and believe that our study will continue to advance our understanding of its applicability for solving real-world problems such as the problem of racial bias in police use of force.

*John Tawa (Mount Holyoke College)*

## Independence FG

### Using Digital Data to Measure and Understand Mental Health

#### SYMPOSIUM

When individuals post to social media or use wearable devices, digital data generated through these everyday interactions with technology reveal a great deal about individuals’ behaviors in ways that were previously not observable. The first talk in this panel shows how social media can be used to better understand the naturalistic manifestations and consequences of loneliness based on a cohort of Twitter users. The second talk shows the potential of patient contributed Facebook data for personalized forecasting of psychotic relapse hospitalizations. The third talk examines the same problem of forecasting relapse through the medium of passive smartphone sensors. The final talk shows explores how social media can be used to characterize two sides of empathy which are increasingly being increasingly linked to stress, excess drinking, missing work and general health. Together, this panel demonstrates the potential for social media and wearable sensors to uncover insight into psychological health and well-being.

*Chair: Sharath Chandra Guntuku (University of Pennsylvania)*

#### Characterizing Loneliness Expressions on Twitter

Loneliness affects approximately 30% of individuals in the United States, and is associated with mental health conditions and high morbidity. We characterized the (online) lives of people who express being lonely and correlate their posts with predictors of mental health. We collected more than 400 million tweets posted by 506,216 users in Pennsylvania, USA, from 2012 to 2015. We identified users who had at least five posts containing the words ‘lonely’ or ‘alone’ (lonely group) and compared them to a control group matched by age, gender, and period of posting. Compared to the control group, users in the lonely group contained more posts about difficult interpersonal relationships, psychosomatic symptoms, substance use, wanting change, unhealthy eating, having troubles with sleep, and in general posted more at night. These posts were also associated with linguistic markers of anger, depression, and anxiety. A random forest model predicted expressions of loneliness online with an accuracy of 77%. Our approach can inform online surveillance for high risk individuals experiencing loneliness and interventions focused on addressing morbidity in this condition.

*Sharath Chandra Guntuku (University of Pennsylvania)*

#### Personalized Forecasting of Psychotic Relapse Hospitalizations with Patient Contributed Social Media

Schizophrenia is associated with significant impairment. While earlier intervention improves outcomes, relapse rates remain high, and limit the established benefit of available services. Existing relapse evaluation strategies such as clinical interview, patient self-report, and family observation are severely limited by their reliance on direct, frequent, and timely contact with trained professionals as well as accurate and insightful patient and family recall. We present preliminary research leveraging patient-generated and patient-contributed

Facebook archives to build and evaluate machine learning methods that predict patients relapse hospitalizations. Specifically, our methods employ both state of the art and innovative natural language techniques and ensemble anomaly detection techniques to forecast relapse hospitalization events in a personalized fashion. We find that given a patient's historical Facebook data, we can predict relapses one month in advance with a specificity of 71% and a positive predictive value of 66%. Our results have implications to advance efforts that capitalize on objective digital data and evidence to improve mental health monitoring.

*Munmun De Choudhury (Georgia Institute of Technology)*

### Forecasting Schizophrenia Relapse Using Passive Sensing

Passively collected smartphone behavioral data present a scalable and at present underutilized opportunity to monitor patients in order to identify possible warning signs of relapse. Seventeen patients with schizophrenia in active treatment at a state mental health clinic in Boston used the Beiwe app on their personal smartphone for up to 3 months. By testing for changes in mobility patterns and social behavior over time as measured through smartphone use, we were able to identify statistically significant anomalies in patient behavior in the days prior to relapse. We found that the rate of behavioral anomalies detected in the 2 weeks prior to relapse was 71% higher than the rate of anomalies during other time periods. Our findings show how passive smartphone data, data collected in the background during regular phone use without active input from the subjects, can provide an unprecedented and detailed view into patient behavior outside the clinic.

*Ian Barnett (University of Pennsylvania)*

### Using Social Media Language to Understand Empathy

The amount and type of empathy that people experience is increasingly being linked to a wide variety of outcomes ranging from social connection to stress and burnout in physicians, therapists and other caregiving professions. We show how social media language, combined with questionnaires, reveals that empathy has both 'good' (compassionate) and 'bad' (depleting) components, with 'bad' empathy associated with stress, reduced perceived control, and reduced well-being, all of which can be measured through people's social media language. These different types of empathy correlate differently with a variety of behavioral health outcomes, including stress, excess drinking, missing work and general health. We also show the utility of a novel annotation methodology in which subjects react to news stories both in free text and in multi-item questionnaire responses.

*Lyle Ungar (University of Pennsylvania)*

## Independence D

### Using Online Technologies to Build Computing Skills That Meet Future Workforce Needs

#### PANEL

Advances in artificial intelligence and the accelerating pace of technology implementation in the workplace have spurred the development of online learning programs in the STEM disciplines, particularly in computer science. Recent interest has highlighted online graduate programs (Goodman, Melkers and Pallais, 2019). Such programs afford working adults the opportunity to build and extend valued job skills while maintaining work and family obligations (Kreth, 2019). Advanced degree programs are also attractive from a national workforce development perspective, where such training may be critical for providing adults (with basic but potentially outdated or insufficient STEM skills) the opportunity to sustain employability, advance their career, or shift occupations. Nonetheless, there remain several critical challenges that confront the success of online graduate programs in computing and their ability to meet future workforce development needs. The purpose of this multidisciplinary, interactive panel is to identify critical gaps and foster dialogue on strategies to build working adult competencies in the technical STEM disciplines.

Understanding and implementing effective online STEM graduate education is dependent on a robust convergence of knowledge across psychological, institutional, and environmental factors. Significant gaps exist in the demand for, experiences in, and impacts of online graduate education, particularly for underrepresented groups in STEM. Online education theoretically offers the potential for diversification (gender/race/ethnicity) of the pool of degree recipients, and therefore future workforce. This discussion will outline the key challenges in this convergence with implications for theory and practice. This interdisciplinary panel is uniquely positioned to offer insights on these issues because it brings these different theoretical perspectives to the table in the on-going study of a large online graduate degree program in Computer Science. Georgia Tech's Online Master's Program in Computer Science (OMSCS) has received national attention and serves as a foundation for discussing the challenges, experiences, and labor market outcomes for adult students pursuing online graduate education in computing.

The Chair/Discussant will introduce the session by describing an interdisciplinary meta-framework that highlights the unique enrollment, motivational, learning, and workforce transition barriers in online graduate learning (10 minutes). As shown in Figure 1, the meta-framework draws from extant knowledge to organize the psychological, sociological, educational, computer science, and public policy influences on adult online learning. Seven panelists will then provide 5-6 minute reviews of evidence and articulate current challenges associated with each thrust in the meta-framework, focusing on the factors that affect the effective and ethical design of learning technologies, how learner attributes, such as age and gender, affect the use of online tools, learning, and career outcomes, key determinants of effective human learning networks, strategies by which learners manage learning within job and non-work constraints, the online technology, and the impact of training on career development. Empirical evidence for the role of these thrusts on learning and career development will be provided using data from the approximately



8000 students who have participated in the ongoing OMSCS program at Georgia Tech. Following the presentations, the Discussant will open the last 15 minutes of the session for questions and comments from all session attendees, and the joint development of a future research agenda.

Members of the proposed panel come from multiple disciplines, including psychology, computing, and public policy. Senior panelists are well-known in their respective fields for their research on adult STEM learning, online technology design, and public policy implications.

*Chair: Julia Melkers (Georgia Institute of Technology)*

### Adult Development and Online Education

Ruth Kanfer will review major findings with respect to age-related changes in cognitive, motivational, and affective influences on learning and what is known about the impact of these changes in online education among individuals with technical knowledge.

*Ruth Kanfer (Georgia Institute of Technology)*

### Institutional Perspectives in the Pursuit of Graduate Education Among Working Adults

Gordon Kingsley will discuss the effects of student motives, institutional environments, and incentives on online enrollment and implications of publicness of work life.

*Gordon Kingsley (Georgia Institute of Technology)*

### Designing Technology for Working Adult Online Education

Daniel Schiff and David Joyner will address determinants and consequences of technology design in online learning. The influence of software developer and technology attributes on design decisions are discussed and implications for the design of ethical and effective digital online systems are described.

*David Joyner (Georgia Institute of Technology)*

### Social Capital and Interpersonal Networks

Isabel Ruthotto will review evidence on the role of social capital and human networks in shaping online educational experiences and outcomes. Informed by social capital theory (Lin, 2001) she will also describe how human networks build and operate in virtual environments.

*Isabel Ruthotto (Georgia Institute of Technology)*

### Going Beyond the Online Lecture

Corey Tatel will review evidence on the role of learner proactivity and discuss methods by which working adults supplement online lectures. Implications for education outcomes and career development are discussed.

*Corey Tatel (Georgia Institute of Technology)*

### Personalizing Online Learning Strategies

Sibley Lyndgaard will use Selection, Optimization, and Compensation Theory (Baltes & Baltes, 1990) to propose how goals and age-related factors may affect online learning strategies and outcomes. Implications for technology design are discussed.

*Sibley Lyndgaard (Georgia Institute of Technology)*

## Independence BC

### Wearable Biometrics Technology: Investigating Its Use in Simulated and Live Law Enforcement Training

#### SYMPOSIUM

This symposium presentation will address the use of wearable biometrics technology and its use in police officer training environments. With the rapid adoption of technology to deliver training, police students can now learn in either live or simulated environments. Our research will consider the physiological differences that are exhibited by police students depending on the training environment. This is an important consideration, especially if the physiological changes and performance in a simulated environment is different than in situ.

Simulation technology has emerged as being an important method used to train new and existing personnel in professions such as medicine, aviation, and policing. This technology allows the user to experience situations that they can expect to encounter in the field, without exposing trainees and/or civilians to unnecessary risks or potential harm. Medical students can practice surgical procedures, pilots can rehearse combat missions, and police can perform emergency response driving procedures, marksmanship practice, and use-of-force decision-making, all in safe and controlled environments (Krätzig, Bell, Groff & Ford, 2010; Reed-Jones, Reed-Jones, Trick, & Vallis, 2007).

Two factors that are known to affect decision-making are physiological arousal and fatigue. The effects of physiological arousal may positively or negatively impact performance depending on arousal level (Martens & Landers, 1970), whereas fatigue typically has a negative effect on performance and decision-making (Vila, Morrison, & Kenney, 2002). Each of these factors are especially common in law enforcement situations. If either of these two factors are experienced beyond an individual's coping ability, it is likely to have a negative influence on task outcomes and decision-making.

Anxiety has been theorized to have an inverted-U relationship with performance and posits that moderate levels of arousal have the greatest positive influence on performance (Martens & Landers, 1970). Low levels of arousal will be too small to initiate the level of effort and attention required to perform a task properly. Conversely, high levels of arousal will be related to high levels of anxiety and stress, negatively influencing performance due to a number of different psychophysiological effects. The purpose of this research will be to develop a greater understanding of the influence of physiological arousal and fatigue on police officer decision-making in police officer training environments. By applying biometric measures of arousal and objective measures of fatigue when investigating police officer decision-making training, we can develop a greater understanding of the influence of physiological arousal and fatigue, alone or in combination, within situations that require quick and effective decisions. This research has important implications for exploring the physiological relationships between live and simulation training and the efficacy of these training modalities to mitigate negative physiological influences, as well as identifying areas in training that may need to be modified. Each of these projects uses a biometric shirt which measures physiological responses such as; heart rate, heart rate variability, and breathing rate to investigate physiological arousal

across four different studies. Additional technology used is functional near infrared spectroscopy (fNIR) which measures cognitive effort (attention, decision-making) through hemodynamic responses.

The first project investigated police performance during a reactive shooter training course. The purpose was to investigate physiological and performance differences during this course in a live and a simulated environment. This project was designed to ascertain if this training could occur in a simulated environment by inducing desired levels of physiological arousal as theorized in live training environments. The second project investigates the efficacy of using driving simulators in emergency response training programs and to compare the arousal levels between driving in simulated versus live vehicles. The third project is an examination of decision-making in a simulated use-of-force scenario while measuring physiological arousal and fatigue. While simulation technology has been used in police training environments for several decades, there is very little research that investigates this as a bonafide training modality. This research has the potential to make policing services more aware of current officer decision-making processes, identify aspects that are most influenced by physiological arousal and fatigue, and provide evidence of simulation technology as an effective training tool. The fourth project investigates decision-making in live, fast-paced and dynamic active shooter scenarios, allowing us to examine how physiological arousal impacts decision-making in these situations. The project will allow us to understand what the effects of physiological arousal are during these situations and what effects this will have on decision-making. These findings will be used to create strategies and training programs to counteract the negative impacts.

Overall, these four projects investigate how non-invasive, wearable technology can be used in police training settings to capture information about physiological arousal in live and simulated environments, and to use these data to make improvements to training. This technology can be used to develop training scenarios to increase desired levels of stress throughout training, understand how decision-making is impacted by stress, and implement strategies to overcome these potentially negative impacts.

*Chair: Gregory Krätzig (University of Regina)*

### **Comparison of Cognitive Demand and Physiological Arousal Between Simulator and Live-Fire Ranges**

Simulation technology has been used by law enforcement for several decades; however, it is less accepted as a training tool for marksmanship training. While there is good evidence that training in a simulated environment provides equal training outcomes (Band, Dragotta, & Sizemore, 2016; Krätzig 2014), there has still been an industry resistance to using this technology. This study was developed to address several logistical problems that included health and safety concerns (e.g., hearing loss and lead contamination), lead remediation, and range access. Other problems that needed to be addressed included cost of ammunition and limited access to ranges affecting training through-put. In an effort to address some of these issues, this study examined simulation as a viable training solution. The training course that is currently being used is a dynamic reactive shooter program that is designed to tax the physical and psychological limits of the student. What was not known was how demanding the course is and if the training could be replicated in a simulated environment. Using 11 experienced police officers who had extensive firearms experience, including being part of Emergency Response

Teams (ERT), Air Marshalls, and Firearms instructors, we trained six of the police officers in the live environment, and five in the simulated environment. The training course was reproduced in the simulated environment and only the baseline and final test were conducted using live-fire, while all other training used laser-based dry-fire pistols. Following consent, all the participants were required to wear a biometric shirt that collected Electrocardiogram (ECG) data. Three of the participants in the live-fire group and three in the simulation trained group also wore the fNIR device that collected hemodynamic data from the pre-frontal cortex; however, connectivity issues in the live trained group resulted in fNIR data being collected from only two of the participants. All participants were tested before the training began to determine their initial skill level. The same test was administered at the end of training to determine if they passed the course, and it also allowed the researchers and instructors to measure performance improvements. The physiological data (ECG and fNIR) will be compared between environments, with event markers being used to measure response differences at specific points in training. The results of the project will demonstrate performance differences between training environments using One-way ANOVA's. Correlations will be conducted to illustrate specific tasks and exercises that provided differences in physiological output as well as those points where differences were not evident. The results of this study may provide more evidence that simulation can be used for firearms training in a dynamic and fast-paced environment and demonstrate the usability of physiological recording equipment to measure performance between environments, testing venues, or course changes. Additionally, significant results of this research will have important financial implications, especially where range access is limited. Simulation can address trainee through-put issues at a fraction of the cost. Simulation also provides realistic training without any of the health and safety concerns, and in most cases time on task is greater in a simulator versus a live range environment.

*Gregory Krätzig (University of Regina)*

### **The Effects of Physiological Arousal During Simulated and Live Emergency Response Driver Training**

As technology becomes increasingly sophisticated, its use in professional training environments continues to grow. Simulation training has been used in aviation for several decades and is proliferating in medical training. However, the use of driving simulators to train police and other first responder drivers is less common. Driving simulation training could be particularly beneficial to first responders, who must perform various dangerous driving tasks in order to respond to an emergency as quickly as possible. Simulated driver training could help prepare first responders to perform some of the most dangerous driving requirements in a safe and controlled environment. Although simulation training may be able to an effective tool for teaching a skill, it is unclear how this technology can prepare users for the increased stress and arousal that will occur during emergency response driving. The effects of physiological arousal on performance have been observed in a number of different environments, including driving. However, the effects of physiological arousal on driving performance are not consistent in research and may vary between simulated and live driving. For example, in simulated driving conditions, it has been observed that increases in physiological arousal are associated with decreases in driving performance (Mehler, Reimer, Coughlin, & Dusek, 2009). In live driving, when confronted with an imminent

crash, Collet Petit, Priez, and Dittmar (2005) found that participants that were more physiologically aroused were more likely to avoid the collision. While other studies have found consistency between live and simulated driving tasks (Reimer & Mehler, 2011). The goal of this research is to compare the effects of physiological arousal on driving performance in both simulated and live training environments with a group of first responders. Data is being collected at a Canadian law enforcement training academy with participants that were currently under going training to become a law enforcement officer. Physiological data will be collected during simulated and live driver training environments in order to understand the individual influences of arousal in simulated and live driving performance, as well as to investigate potential differences in arousal and unique effects between these environments. Physiological arousal data will be collected using smart shirts that participants wear under their normal attire. This shirt has two straps, one across their thorax and one across their abdomen, that secure sensors against the participant, which will measure Electrocardiogram (ECG) and breathing rate. Simulated driving training involves responding to a call and navigating through four red-light intersections. This urban drive requires participants to follow procedures for driving through each of the red-light intersections, sometimes while traveling in oncoming traffic lanes. During the drive participants are also required to communicate with a dispatcher who will be providing details about the emergency and asking the participant to respond to questions about their location and estimated time of arrival. Live driver training is measured in two contexts to capture a full picture of arousal and live driving. The first context requires responding to an emergency in a live track within the training academy. However, some participants' driving scenario will involve a pursuit, while others' will not. The participants' performance will be measured on their ability to complete the manoeuvres on the track, as well as on their communication skills and decision-making throughout the scenario. The second context of live driving requires participants to perform various driving procedures within an urban environment. Within this driving context, participants' performance will be based on their ability to perform a number of different driving procedures, including speed maintenance, U-turns, and lane changing. The data for the simulated driving scenario have been collected; however, data in the live environment is scheduled for collection between March and June 2019. The analyses for these data will focus on predicting driving performance with physiological data using regressions. Direct comparisons will be made between physiological arousal and observed changes in arousal between simulated and live environments. Indirect comparisons of the associations of physiological arousal and driving performance will also be made between simulated and live training environments. The results of this study have important implications for understanding emergency response driver training for first responders, including law enforcement, firefighters, and paramedics. These results will be able to demonstrate whether arousal has a similar effect on emergency response driving or if the training received prepares first responders to manage arousal under these conditions. Important implications will also be made about the efficacy of simulated driver training as an effective, analogous method for training emergency response driving. Lastly, these results will evaluate the need for training programs to incorporate aspects of stress and arousal into their training programs in order to ensure that first responders are able to manage the effects of arousal and safely arrive to the emergency.

*Chet Hembroff (University of Regina)*

### **How Arousal and Fatigue Influence Decision-Making in a Simulated Use-of-Force Training Environment**

Police officers interact with the public in many ways, from minor inconsequential interactions where an individual asks for directions to large complex interactions where an officer is de-escalating a dangerous situation. Each situation is influenced by factors determining how the officer and civilian(s) will interact; individual aspects of self-care and experience, environmental aspects of uncertainty, and psychological and physiological aspects of stress, anxiety, cognitive effort, and fatigue. The project investigates these factors and how they impact community-police interactions. The research compares policing students who are responsible for identifying a response option in a simulated use-of-force situation and those responsible for implementing a response option during the same situation, and whether conditions such as fatigue, physiological arousal, and cognitive load influence decision-making ability and performance. This has the potential to reinforce the necessity of scenario-based training, wherein students can learn and practice assessing situations while rapidly identifying a course of action, and implementing this action. The implementation phase in scenario-based training can expose trainees to varying consequences of their actions which results in growing their mental library and builds upon their development of experience-based decision-making. High levels of stress are correlated with risky decisions, attention to threat-related material, and increases focus on task-irrelevant stimuli, resulting in individuals making the wrong decision or missing cues that may result in an alternative or better course of action (Akinola & Mendes, 2012; Nibbeling, Oudejans, Ubink, & Daanen, 2014; Nieuwenhuys & Oudejans, 2011; Pabst, Brand, & Wolf, 2013). Police officers are also at a higher risk for sleep-related issues as they often work shift work, in which less sleep is associated with increased difficulty in remaining attentive and responding in a timely manner. Therefore, the interaction between stress, fatigue, and decision-making is a vital area of research because even if an officer improves one factor, deficits in other areas can severely impact overall decision-making performance. Data will be collected with law enforcement trainees through March and April 2019 and analyzed in June 2019. Participants will be required to wear a smart shirt that will measure Electrocardiogram (ECG) data. Other data collected will be demographic information, and scores on several psychometric tests of decision-making, stress, and fatigue. Participants will also be required to either watch or interact with a simulated use-of-force scenario and complete a post-scenario interview. Decision-making and fatigue are measured through two computer-based tests; Iowa Gambling Task (IGT) and Psychomotor Vigilance Test (PVT). The IGT measures risky and impulsive decision-making by modeling real life decision-making including factors of uncertainty, reward, and punishment (Barnes, Gunia, & Wagner, 2014; Killgore, 2010). The PVT measures alertness and vigilant attention (Basner & Dinges, 2013; Killgore, 2010). The simulated use-of-force scenario will be shown to Identification Group participants (scenario viewing only) and at the time of decision-making, the video will pause, and the participant must verbalize the response option they would implement if they were in the situation. The Implementation Group (scenario interaction) will interact with the simulation and at the decision point, they will be required to de-escalate the situation and/or stop the threatening individual from

causing harm to themselves or others. The interview will require the participant (in both groups) to identify and explain their choice and alternate options that may be successful during this situation. The proposed project intends to investigate the interactions of the measurements and factors, including both main effects and interactions. Therefore, the proposed project will include a multivariate analysis of variance (MANOVA) of the factors. Subsequently, Bonferroni-corrected ANOVAs will be conducted to determine which variables are significantly different. Overall, this research has the potential to make policing services more aware of current officer decision-making processes and techniques and identify which aspects are most influenced by stress and fatigue. Additionally, this research will serve as a validation of the practicality of wearable biometrics in dynamic police officer training environments and investigate how much stress is induced in simulated judgment situations. Thus, ensuring Canada's training of police officers is evidence-based and equipping them with the appropriate strategies and techniques for reducing stress and fatigue related errors.

*Tansi Summerfield (University of Regina)*

### **The Impact of Arousal on Skilled Performance in High-Stress Dynamic Training Environments**

Police officers often have to make decisions in high-stress dynamic encounters, and it is important to ensure that scenario-based training closely resembles real world situations. This project investigates decision-making in a physiologically arousing, dynamic, immediate-action live scenario. In recent years there have been an increase in active shooter situations in both the USA and Canada, and in response to this many police forces have developed training programs to help train recruits and officers to help ensure the safety and security of citizens. This program is often referred to active shooter training and is best defined as the swift and immediate deployment of law enforcement resources to an ongoing, life threatening situation where delayed deployment could result in death or grievous bodily harm to innocent persons. Active shooter training was designed to allow police to engage a threat immediately and take the necessary action to contain the threat. In order to successfully complete this training, police require skilled performance in which skilled performance refers specifically to the performance of motor-skill ability. This is a three-phased process of skill acquisition and is learned through training and practice (Fitts & Posner, 1967). First, specific task demands are learned, then these skills are refined, and finally mastery of skills is learned and perfected (Chung et al., 2005; Fitts & Posner, 1967; Krätzig, Hyde, & Parker, 2011). For law enforcement firearms training this process refers to learning tasks such as shooting from different positions (e.g., standing or prone), tactical breathing (e.g., slowing heart-rate), trigger pull, and grip. These skills are refined through repeated practice of drills, and finally mastery is demonstrated by successfully completing a formal evaluation (Chung et al., 2005; Fitts & Posner, 1967; Krätzig, Hyde, & Parker, 2011). This skill development theory posits that the third phase (mastery) results in performance automaticity wherein the individual no longer needs to focus their attention on how to do a particular task. It is at this stage that the high stress active shooter training is introduced with the expectation that individuals have acquired the necessary skills to successfully accomplish this task. Stress is a maladaptive state where the body undergoes multiple automatic physiological responses in an attempt to restore homeostasis (Pabst, Brand, & Wolf, 2013; Ross & Vilke, 2017; Starcke, Wolf, Markowitsch, & Brand, 2008). This response

occurs through two main pathways; one pathway is fast acting (nervous system) and involves the activation of the sympathetic nervous system, which can result in an increase in heart rate, and breathing rate (American Psychological Association [APA,] n.d.; Pabst et al., 2013; Ross & Vilke, 2017). The second pathway is slow acting (endocrine system; APA, n.d.; Pabst et al., 2013; Ross & Vilke, 2017) and involves the hypothalamic-pituitary-adrenal axis (HPA-axis), which can result in a shift to more intuitive decision-making (APA, n.d.; Pabst et al., 2013; Ross & Vilke, 2017). This stress response directly and significantly impacts decision-making (Pabst et al., 2013; Ross & Vilke, 2017; Starcke et al., 2008). Research has found that individuals make riskier and more disadvantageous decisions when stressed (Pabst et al., 2013; Starcke et al., 2008), and these decisions tend to be less conscious, less deliberate, and less analytical when under increased stress levels (Ross & Vilke, 2017). These physiological responses to stress make it important to investigate how high-stress active shooter training is experienced. Participants will be required to wear a smart shirt that will measure physiological arousal by collecting electrocardiogram (ECG) data during their training. Initial data has been collected to ensure that the smart shirt records high quality readings during this training. Further data collection will take place in May, 2019 and will be analyzed in Summer, 2019. The proposed project intends to investigate the level of physiological arousal that is experienced during active shooter training, and its impact on decision-making. How physiological arousal impacts performance specifically regarding the prevalence of common errors (e.g., clearing corners, missed shots) will be measured. This research has the potential to help improve the active shooter training to ensure more effective decision making and ensure it is eliciting physiological arousal to simulate realistic active shooter situations. Thus, leading to possible strategies to counteract the negative impacts of stress on skilled performance, and improve public safety.

*Billea Ahlgrim (University of Regina)*

---

## **Independence HI**

### **Developing Technology-Based Mental Health Interventions**

#### **PAPER SESSION**

#### **An Online Study On the Evaluation of Dr. Babel Fish: An Unsupervised Machine Translation System for Simplifying Providers' Language**

The healthcare community increasingly uses patient-centered and in-home care models, requiring patients to access and understand a large amount of health information [2,3]. Much of this information is web-based. Most notably, patient portals have become widely adopted, allowing patients access to their Electronic Health Records (EHR) [4]. However, benefits of patient-centered care can be hindered by poorly designed information, [4] especially for patients with low levels of health literacy [7].

This mismatch in patient literacy levels and health documents is due in part to the mismatch in preferred medical terms by healthcare professionals and patients [8,9]. For example, what professionals refer to as "abdominal pain", patients might refer to as "stomach

ache". We aim to bridge this gap by developing an unsupervised machine translation system (named Dr. Babel Fish; demo at <http://drbatisfishweb.mybluemix.net/>) to translate complicated professional statements to their patient-friendly counterpart, and we take medication instructions present in EHR as a use case. While pharmacists traditionally assume part of the responsibility for translating technical e-prescription information to patient instructions, our system can reduce the burden on these professionals.

Dr. Babel Fish is designed as a 3-stage pipeline (see Figure 1) inspired by Phrase-Based Machine Translation systems [6]. In the first stage, complicated/hard (and easy) words are identified based on frequency of usage in a patient-friendly medical corpus. Next, possible translations of a given complicated source word are collected and each given a translation probability. We use the Unified Medical Language System [1]: a repository of biomedical words/phrases, to gather the translations and estimate their probabilities. Finally, every candidate output translation is assigned a language model score and a translation model score. The highest scoring translation is then selected as the output.

Previous work addressing this problem performs word replacement using dictionaries and ontologies, but ignores the issue of ambiguity [5,10]. For example, a term like "mg" in the original text might either mean "Milligrams" or "Magnesium". By utilizing an unsupervised machine translation framework and its language model component, we address the aforementioned issue of ambiguity, which is of necessity in this medically critical setting. We compare our system to strong baselines lacking a language model, and observe that our system achieves the highest level of accuracy at 88% vs. the second best at 79.48%.

Finally, we performed an online study to evaluate the impact of our system's translation on patient comprehension. One hundred medication instructions were randomly extracted from an EHR system and served as input to our translation system (e.g., original untranslated: "Takes 0.5 ml bid prn"; translated: "Takes 0.5 ml twice a day as needed"). 160 participants (52 older and 108 younger; mean age = 45.1; range = 18-79; 89.4% greater than HS degree) saw an equal number of untranslated and translated instructions with three different levels of hardness, defined as the numbers of words identified as hard by our translation system. For each instruction, comprehension was measured by accuracy in answering multiple-choice questions.

We conducted two analyses. The first included all 100 sentence pairs to evaluate the effectiveness of the algorithm. The second only included successfully translated pairs to evaluate benefit of translation for comprehension.

All sentences: Comprehension accuracy was analyzed by a 2 (translated/untranslated) x 2 (age: older:  $\geq 60$  and younger) x 3 (hardness levels) mixed design ANOVA (condition and hardness were repeated measures). Comprehension was better for sentences that were translated ( $F(1,7973)=112.3$ ,  $p < .0001$ , 58% vs 47%) and with hardness levels 1-2 versus 3 ( $F(2,7973)=14.6$ ,  $p < .0001$ , level 3: 49% < levels 1: 55% and 2: 56%). Notably, older adults better understood the sentences ( $F(1,7973)=224.4$ ,  $p < .0001$ , 64% vs 47%). As Figure 2 (top graph) shows, the effect of translation depended on hardness level ( $F(2,7973)=14.0$ ,  $p < .001$ ).

Successfully translated sentences: Two clinicians identified successfully translated sentences, (89% agreement). As before, comprehension was better for translated sentences ( $F(1,5108)=168.0$ ,  $p < .0001$ , 63% vs 46%), although it was not influenced by hardness

( $F(2,5108)=0.4$ , ns). Older adults again better understood the sentences ( $F(1,5108)=169.2$ ,  $p < .0001$ , 67% vs 49%). Also as before, the effect of translation depended on hardness level ( $F(2,5108)=7.6$ ,  $p < .001$ ; see Figure 2, bottom graph). Notably, older and younger adults benefited equally from translation (Age x Translation  $F(1,5108)=0.3$ , ns).

Dr. Babel Fish not only outperformed alternative baseline approaches, but increased patient comprehension of medication information. Older adults better understood the medication information, presumably reflecting more experience with and interest in health information, and benefited as well as younger adults from translation. Nevertheless, increasing the accuracy of translation is an important goal, as well as human-in-the-loop integration of the system to ensure 100% accuracy output. Next steps also include integrating Dr. Babel Fish with a computer agent-based health adviser system, where translated medication information is presented by an embodied agent in order to engage patients with their self-care.

*Tarek Sakakini (University of Illinois at Urbana-Champaign)*

### A Novel Approach to the Evaluation of mHealth Apps

**Problem Statement:** Mobile apps are being used in health care to support prevention efforts, as well as patient assessment, treatment and self-management. While these apps are easily accessible and often free patient or provider resources, the evidence behind some apps is lacking. To reduce the potential for misinformation, false diagnosis, and harmful interventions, patients and providers need guidance on which apps to use. The evaluation of mHealth apps with a structured rating tool provides users with systematic and objective information to support the informed use of these technologies. While existing app rating tools such as "MARS" and "PsyberGuide" reliably assess mobile apps' usability (Stoyanov, 2015) and evidence backing (Neary 2018), the Defense Health Agency (DHA) Connected Health team felt a gap still existed on the app rating market. There was not a single tool that had objective measures, multiple construct areas, and could be used for all physical and behavioral health conditions.

DHA Connected Health created an objective app rating tool to inform mHealth app selection for DOD Beneficiaries. The following is a use case example, highlighting the value and implications of the App Rating Inventory (ARI).

**Procedures/Analysis:** A search of Google Play (Android) and App Store (iOS) was conducted using the search term "Autism". An initial search result of 250 mobile apps was funneled down to 19 apps based on the following inclusion criteria: patient-focused, child specific, involve a treatment approach based on applied behavioral analysis (ABA), include education and treatment support. These apps were further narrowed to 10 using publically available user generated data (number of user reviews, ratings, and downloads). The 10 apps with the highest user generated data were then rated along three dimensions (evidence, content, and customizability) with the 28-item ARI. Five apps were removed from the app distribution platforms prior to rating. Three of the remaining five apps met or exceeded a 50% agreement threshold (scoring at least 14 or higher on the 28-item inventory) and are shown above the red line in Table 1.

**Implications/Conclusions:** The ARI was designed with the assumption that clinicians will be most interested in apps that are evidence-based, have good content, and are easy to use. Note, though, that higher scores on the tool do not necessarily indicate that an app is superior or preferable; in fact, a stakeholder may determine

that an app that is especially customizable is preferred over an app with a moderate evidence score and would orient their app selections accordingly.

In our example, “Autism Therapy with MITA” obtained the highest possible score on the evidence scale. This application has a firm evidence base, including an RCT. While “Otismo-Special Education ABA Therapy Autism Game” and “Autism Early Intervention App” scores are nearly perfect for evidence base, the apps have not been included in a formal study. “ABA Therapy Aphasia Autism” and “Autism How to Help your Child” have also not been studied.

There was even more distinction between the top five apps on the content scale. “Autism Therapy with MITA,” “Otismo-Special Education ABA Therapy Autism Game,” and “Autism Early Intervention App” scored higher than the other two apps on treatment-focus items, obtaining a positive rating for providing information regarding types of treatments that are commonly administered to reduce the problem occurrence, and providing skills-learning features to help ameliorate symptoms. “ABA Therapy Aphasia Autism” and “Autism How to Help your Child” did not receive a positive rating as they did not include interactive content.

Along with high evidence scores, “Autism Therapy with MITA” and “Otismo—Special Education ABA Therapy Autism Game” received the highest scores for content and customizability.

It is important to note that all of the apps identified are intended to supplement the home life of children enrolled in ABA treatment. These apps are not intended to provide standalone treatment to children with autism.

The DHA Connected Health ARI has the potential to impact active duty military and their families, as well as civilians by providing objective ratings of mobile apps that may be used as part of patient care.

The ARI app rating tool has not yet been published. The first round of analysis on the DHA Connected Health app rating tool has been completed. This stage included content validity, pre-testing of objective measures, sampling a variety of mobile apps, factor analysis and item reduction. The second round of analysis is in progress testing the dimensionality of the tool, reliability across topic areas, and continuing interrater reliability within the app rating team (Boateng, 2018).

*Rachel Mackey (DHA Connected Health)*

### **Crowdsourcing Stressors: Designing Digital Mental Health Apps to Reflect Community and Support the Therapeutic Process**

The first encounter with a digital mental health app, much like a first encounter with a skilled clinician or counselor, should exemplify an understanding of patient needs and their situation to promote therapeutic alliance and support engagement. With the development and use of digital mental health apps on the rise, a growing community of researchers, practitioners, and users are defining how to design, evaluate and implement digital mental health apps to best support patients (Toros et al, 2018). Unfortunately, the utilization of such apps is low and the majority of apps are designed for monitoring symptoms and patient education. They position the patient as a passive data stream to be evaluated or an audience for self-motivated education. They are useful yet remain detached from the patient's relationship with the provider and limit patient expression of their condition. By directly engaging communities of need in the design process, we can design mobile health apps that accurately reflect

individual and community needs as a step toward improving engagement and integration of digital mental health apps into the therapeutic process.

In this talk, we present the design and development of a tablet-based, digital mental health app, called Life Map, designed to support college students during their first experience in one-to-one psychotherapy with a counselor or therapist. The app helps students identify their stressors and elect focus areas for their therapy sessions. The app was designed using participatory methods engaging over 300 students, counselors, and mental health subject matter experts from one university and a local health care system. In this talk, special attention will be given to the evolution of the knowledge representation of stressors in the app that was directly crowdsourced from the college community and subject matter experts in focus groups and validated using the saturation method. The resulting set of stressors balances the student expression of their needs and representation of clinical knowledge. Ultimately, the list of 93 stressors reflect the stress of college students at a specific university at a specific time (e.g. “getting enough sleep,” “accessing health services,” “caring for pets,” “sadness,” “using caffeine or stimulants,” “identity” etc.). The stressors function in the app as a way for students to express their emotions, goals, and quickly build rapport with their therapists. The collection of stressors from a broad range of students can also assist counseling services to create responsive programming based on local student population trends rather than national trends or end-of-semester surveys which lack specificity of the immediate community and suffer from delayed implementation.

In an observational study of the app, students created a Life Map in the waiting room of counseling services before each therapy session to express their current situation and stress, and then used the app to support communication and decision-making with their therapist during their session. We collected data from 53 students immediately prior to one-on-one counseling sessions using the “Life Map” digital mental health mobile app. A total of 203 “maps” of stressor data were collected. Figure 1 shows the app as viewed by a single student showing the stressor tags selected and their orientation on the map. Figure 2 shows the aggregate of tags collected over the course of the study. Surveys and post-session interviews with providers were assessed to identify strengths and weaknesses of the app in facilitating communication, expanding or focusing the session, and therapeutic value. The focal point of this talk will be the participatory design process and knowledge representation; however, it is worth noting that statistical analysis of the study data shows a strong correlation between the Life Map tag placement and PHQ-9 scores and high-acceptance by users ([Blinded], in preparation).

As knowledge from patients with mental health conditions is elicited and captured by technologies and health care systems, either actively or passively, design decisions about how people are represented and how their needs are expressed have implications for access, adherence, and quality of care.

*Barbara Barry (Mayo Clinic)*

---

## **Independence E**

### **Robotics and Psychology**

## PAPER SESSION

**Systematic and Multidimensional Perception of Robot Minds**

Whether robots have minds is partly a philosophical question, partly an engineering question—the answer depends both on one’s conception of “mind” and on the robot’s actual capacities. Whether people treat robots as having minds is largely a psychological question—its answer depends primarily on people’s expectations, perceptions, and inferences, even if the robot’s actual capacities are quite different. This predominant impact of perceptions is due to people’s evolved tendency to treat nonhuman entities as having agency and mind if they display a few critical features: having eyes, contingently responding to the observer, or showing self-propelled, equifinal behavior (Johnson, Slaughter, & Carey, 1998; Luo & Choi, 2013). Even nonbiologically moving objects (fury blobs, boxes, or triangles) are treated as if they have intentions and emotions when they exhibit continuous, goal-directed movement (Heider & Simmel, 1944; Király, Jovanovic, Prinz, Aschersleben, & Gergely, 2003).

This tendency to anthropomorphize (Epley, Waytz, & Cacioppo, 2007) is not shallow or unidimensional, however; people do not indiscriminately attribute “more mind” the more triggering features an entity shows. But just how many dimensions of mind attribution are there? And how many do people grant robots? Some researchers have proposed six categories of mind (D’Andrade, 1987), but these may reduce to two (Haslam, Kashima, Loughnan, Shi, & Suitner, 2008), defined by perception and cognition-vs.-emotion. Others also proposed two dimensions, but different ones, namely experience and agency (Gray, Gray, & Wegner, 2007). Yet others proposed three dimensions, distinguishing emotion, intention, and cognition (Kozak, Marsh, & Wegner, 2006), or bodily experiences, emotions, and perception/cognition (Weisman, Dweck, & Markman, 2017). Despite these variations, a common observation is that people are reluctant to attribute capacities along an emotion/experience dimension to robots, whereas they typically grant them some capacities along a perception/cognition/intention dimension (Gray et al., 2007; Malle & Thapa Magar, 2017; Sytsma & Machery, 2010; Weisman et al., 2017). However, this observation demands refinement: Do people see no emotions at all in robots? Are there no further differentiations among the capacities they grant robots? And do these patterns hold across different robots? In a series of studies we garnered three new insights into people’s conception of mind in general and their perception of mind in robots in particular.

(1) Because previous studies differed widely in the specific capacities they examined, we collated a pool of over 70 human mental capacities and showed in four studies that they separate into three consistent dimensions of mind, in both humans and robots: Affect unites aspects of physiological and emotional capacities; Moral and Social Cognition combines moral regulation with the simulation of one’s own mind (e.g., planning) and others’ minds (e.g., inferring their thoughts); Reality Interaction represents dynamic agent-world interactions from perception and cognition through learning to action and communication (see Figure 1 for high-loading items in one recent study).

(2) By varying the specific judgments people made—what capacities they believe a robot to have vs. what capacities they would like a robot to have—we found that two of the above dimensions can come apart: Affect divides into Positive and Negative Affect; Moral-Social Cognition divides into Moral and Social Cognition. Thus, people’s

conception of mind may be organized in a potential five-dimensional space in which some dimensions merge depending on the judgment task or the objects of judgment. From these results we developed a 20-item measurement instrument that contains reliable 4-item subscales for each of the five dimensions, which can also be combined to yield scales for the three primary dimensions (see Figure 2).

(3) Using this instrument, we have begun to examine mind attribution patterns across different robots and found that specific dimensions are differently responsive to robots’ appearance, function, and role. For example, when considering what they would like to see in a robot, people were open to Positive affect (e.g., compassion) in a home robot but not a military robot, though in both cases they strongly rejected Negative affective capacities (see Figure 3). As another example, in an ongoing study presenting people with videos of various robots, we find that all five dimensions discriminate between of a robot with low humanlike appearance and robots of high humanlike appearance; but only Moral cognition discriminates between the two robots of high humanlike appearance (see Figure 4).

In sum, our studies suggest that people do not blindly anthropomorphize robots; on the contrary, they systematically infer specific mental capacities from a robot’s form and function. With these insights, and equipped with a reliable measure of the three- to five-dimensional structure of mind perception, researchers can tackle many new investigations—into developmental, cross-cultural, and longitudinal dynamics of human-robot interaction.

*Bertram F. Malle (Brown University)*

### Developing an Affordable Robot Companion for Elderly Support

An aging America is challenging the U.S. healthcare system. As the baby-boom generation approaches retirement, each passing year increases the pressure on healthcare economies. Current and future professional care staff will be unable to treat or even monitor the wide-ranging challenges of the aging population—from anxiety to loneliness, from dementia to physical disability (Eden, Maslow, Le, & Blazer, 2012; World Health Organization, 2013). Individual citizens and the health-care system as a whole urgently need new solutions to these accelerating problems.

One way to alleviate these pressures is to use intelligent technology to augment the system of care. Technology will not, and should not, replace trained clinicians, health-care workers, and certainly not family members. Instead, technology can offload some of the responsibilities of overburdened family members and of an overtaxed healthcare system, thereby providing more care to more people. The right kind of human-centered technology can even re-engage older adults to participate more actively in their own lives despite challenges and deficits. But to be human-centered, such technology must be designed with a detailed understanding of older adults' challenges in daily living and the prospects of technology to offer acceptable and trustworthy assistance. Smartphones and their apps are not human-centered in this way. Such devices are powerful but also opaque. They demand that the user adapt to the system, not the other way around—an ill-advised feature in technology for older adults.

In contrast to disembodied technology, physically present artificial agents—"robots"—have a variety of advantages for attention, motivation, and enjoyment (Bainbridge, Hart, Kim, & Scassellati, 2011; Fasola & Matarić, 2013; Kiesler, Powers, Fussell, & Torrey, 2008). Physical presence and signs of agency, even in simple robots, trigger curiosity, engagement, and activity (Takayanagi, Kirita, & Shibata, 2014); these benefits increase over time (Šabanović, Bennett, Chang, & Huber, 2013); and interacting with a small robot can reduce depression among patients with dementia (Jøranson, Pedersen, Rokstad, & Ihlebæk, 2015).

In the present project we are collaborating with an industry partner to develop a small robot that is affordable, provides physical and psychological comfort due to its zoomorphic form, and has limited but focused intelligent capacities, such as helping a person track misplaced objects (e.g., reading glasses, keys, etc.). In performing its tasks, moreover, the robot is not a servant but collaborates with the care recipient.

We report here on the first phase of an ongoing study (N = 50 adults over 65 years old) and will have substantially more data within a few weeks (data of 100 adults over 65 years old and 100 informal caregivers of older adults have been collected and will be analyzed shortly).

The first sample was recruited primarily from independent living facilities, where participants completed standardized surveys in an in-person administration. Average age was 81.4 (67 to 95), with a strong gender asymmetry of 84% women. Most participants reported having some memory difficulties (69% "a little bit," 10% "quite a bit"), but only one had been diagnosed by her doctor as having dementia. We focus here on two major results emerging from this sample.

We asked participants to evaluate 21 challenges of daily living (partially derived from...), including housework, technology,

finances, and so on (see Table 1 for full list). People indicated to what extent they experienced each difficulty (0-No to 5-Almost Always). Keeping in mind that this sample is highly functional in most domains, we see low ratings overall (mean of 0.3), but the top five difficulties stand out. Technology (M = 1.2), misplacing or losing things (M = 1.0), moving oneself (M = 0.8), moods (M = 0.6), and speech and language (M = 0.5). Within our project, this result provides initial support for two of our assumptions: (a) that the robot needs to be accessible, not provide another technical obstacle; and (b) that helping with misplaced objects is an important function to implement in the pet companion.

We also asked participants for which of 12 tasks they would adopt (0-Not likely to 2-Very likely) a companion robot (see Table 2 for full list). The top tasks were measuring vitals (M = 1.0), followed by locating lost objects (M = 0.7), fall detection (M = 0.7), playing cognitive games (M = 0.6), medication reminders (M = 0.6), and connecting with friends and family (M = 0.6).

All of these tasks are technically feasible and within our robot's planned capacities. The biggest technical challenge currently is for the robot to track the location of misplaced objects and use intuitive nonverbal communication to guide the user to the objects. By October we will have at least a prototype solution to present.

*Bertram F. Malle (Brown University)*

### Investigating the Cross-Category Effect of Face Perception in Human Robot Interaction

Our ability to process social information, such as facial identity, contained in human faces allows us to successfully engage in human-human interactions. As interactions with robots become more and more social and ubiquitous, it is also important to consider face perception in the context of human-robot interaction (HRI). Engineers are working hard to make robots appear more physically human-like since perceiving robots as entities with human-like features improves performance during social interactions (Kiesler, Powers, Fussell, & Torrey, 2008; Özdem et al., 2017; Scassellati, Admoni, & Matarić, 2012). As rapid technological advances eventually lead to robots that are so realistic that they are visually indistinguishable from humans, it is important to examine to what extent these robots trigger processes relevant to social cognition, such as face perception. Previous research suggests that whether interaction partners are considered a part of a perceiver's own ingroup modulates social cognitive processing, with positive effects for ingroup faces and negative effects for outgroup faces (i.e., cross-category effect; Bernstein, Young, & Hugenberg, 2007). Most importantly, group membership modulates social cognition even when controlling for physical appearance (i.e., identical stimuli), and manipulating membership status merely by instructing participants that human stimuli represent members of a social ingroup (i.e., minimal groups; Brewer, 1979) or that nonhuman stimuli "have a mind" (i.e., mind-face link; Deska & Hugenberg, 2017). These studies suggest that basic mechanisms of face perception, such as configural processing (i.e., perceiving a face stimulus holistically), might be influenced by top-down modulatory effects related to beliefs regarding an agent's human-likeness. Configural processing is traditionally measured via the face-inversion effect, such that presenting a stimulus upside-down as opposed to upright reduces recognition performance for face-like stimuli due to a stronger disruption of configural processing but not for nonface-like stimuli. The difference in performance with upright faces compared to inverted

faces represents the inversion effect and is a measure of the extent of face-typical processing (i.e., larger inversion effects indicating more face-typical processing; Farah, Tanaka, & Drain, 1995). However, while the CCE and mind-face link has linked group membership and mind status to stronger face-typical configural processing, this has not been looked at in the context of HRI.

The goal of the current experiment is to examine whether beliefs about the human nature (i.e., human versus android) of an agent changes how face-like a stimulus is perceived that actually represents a “real” human. For this purpose, we use images of humans, embedded in a face-inversion paradigm, and instruct participants in one condition that they represent human agents and in the other condition that they represent android robots that are made to look identical to humans. To make the android manipulation believable, we showed participants images of existing android robots that are hard to distinguish from humans when they do not move (e.g., “ISHIGURO Symbiotic Human-Robot Interaction Project,”), and provided them with a background story explaining that were working with an android designer who designed robots that were nearly indistinguishable from humans and that we were interested in their perceptions of them. Since robots are perceived as having less mind than human agents (Wiese, Metta, & Wykowska, 2017) and are not part of the “human-ingroup”, we predicted configural processing (i.e., measured via the face inversion effect) to be less pronounced in the android versus the human condition. Stimuli consisted of 80 greyscale images of white male faces that had an equal likelihood of appearing in the human or android condition. 48 participants were recruited via Mechanical Turk for the study. The experiment consisted of two parts: during the learning task, participants passively viewed 20 faces (3500 ms each) that were presented with a colored frame (green for human, blue for android) and a ‘Human’/‘Android’ label. Participants then advanced to the recognition task, in which the 20 faces from the learning task and 20 new faces (with half of the original and half of the new faces being inverted) were presented, and participants had to indicate via key press whether they had seen a given face during the learning task. Hit rates and false alarms were calculated separately for upright and inverted faces for humans and androids and used to calculate sensitivity ( $d'$ ) scores.  $D'$  scores were entered into a 2 (Agent Belief: Human, Android) x 2 (Orientation: Upright, Inverted) repeated-measures ANOVA.

The results showed that participants had stronger inversion effects when they believed the agents represented humans rather than androids; see Figure 1, providing more evidence that an agent’s social category (in line with Bernstein et al., 2007), as well as mind status (in line with Deska & Hugenberg, 2017) can influence face-typical processing. Our results also suggest that simply designing a robot that is visually indistinguishable from humans is not enough to fully engage social-cognitive mechanisms important for HRI, and that other methods should be considered to accomplish this.

*Ali Momen (George Mason University)*

### **Robopsychology: Towards a Convergence of Deep Learning Artificial Intelligence and the Science of Mind and Behavior**

The past decade has seen remarkable advances in many artificial intelligence applications, driven in large part by deep neural networks—a branch of machine learning that uses simulations of the brain to learn a mapping between inputs and outputs. Deep learning arguably represents the most promising computational theory of the brain to date, delivering human-level performance on many real world tasks. However, despite its origins in psychology and neuroscience, little recent work has considered deep learning as a model of biological intelligent behavior. Here, we introduce an integrative approach to develop and test theories of behavior and the brain by embedding deep neural networks in a behaving robot. The robot is connected via wifi by a high-powered GPU computer to which it sends its camera images. The computer sends the images through a deep neural network and receives the network’s action decisions. This approach allows for the co-development of behavioral and neural theories and a means of testing them in a novel ‘model organism’ whose brain can be experimentally manipulated in a manner that is impossible in a biological organism. In our research, we have used this platform to test numerous theories of the brain during supervised and reinforcement learning tasks. These include models of compression of perceptual data, the role of development, attention, and curiosity in accelerating learning and comprehensive comparisons of performance across multiple deep learning architectures. Overall, we believe this platform and approach demonstrates the potential for synergistic research across behavioral, computational and neural approaches for understanding the brain and behavior while reciprocally improving approaches to applied machine learning.

*Elan Barenholtz (Florida Atlantic University)*

12:00–1:15 P.M.

## Lunch On Your Own

---

1:15–2:30 P.M.

## Concurrent Sessions 2

### Independence D

#### Interactivity with Automation

##### PAPER SESSION

#### **Your Human-Likeness Matters to Me: Is Categorical Perception a Potential Cause for the Uncanny Valley?**

Human-like but not perfectly human appearing agents frequently evoke feelings of discomfort and eeriness, a phenomenon termed the Uncanny Valley (UV; Mori, 1970; Kaetsyri et al., 2015). The Categorical Perception Hypothesis proposes that effects associated with the UV might be due to uncertainty as to whether to categorize agents falling into the valley as “human” or “nonhuman” (Cheetham et al., 2011), and empirical data has shown that being exposed to uncanny agents leads to a depletion of cognitive resources over time, even when the agent’s human-likeness is task-irrelevant (Wiese et al., 2018). Although empirical evidence in support of the UV is increasing, there is no consensus yet regarding its theoretical underpinnings (see Kätsyri et al., 2015; for a review). One promising hypothesis purports that the physical appearance of humanoid agents triggers a categorization-related cognitive conflict as to whether the agents represent human or nonhuman entities, and that this conflict results in negative emotional evaluations due to increased cognitive processing costs needed to resolve categorical ambiguities (Categorical Perception Hypothesis; Cheetham et al., 2011), which is in line with insights from evolutionary biology that link categorization to survival, and the failure to categorize certain stimuli to negative emotional responses (Burleigh & Schoenherr, 2015). The aim of the current experiment is to examine to what extent typical phenomena associated with the UV (e.g., increased cognitive conflict processing) are specific to categorical uncertainty regarding agents’ human-likeness as opposed to being related to categorical ambiguity in general (i.e., category A vs. B).

In the current experiment (165 participants), we used mouse tracking to determine whether agent spectra with (i.e., robot-human) and without (i.e., robot-animal and robot-stuffed animal) a human endpoint cause categorical perception to comparable extents. Specifically, we compared human and nonhuman agent spectra with respect to the existence and location of a category boundary (H1-1 and H2-1), as well as the magnitude of cognitive conflict around the boundary as measured via maximum deviation of mouse movements (H1-2 and H2-2). The results show that (a) both human and nonhuman spectra exhibit category boundaries (H1-1) at which cognitive conflict processing is higher than for less ambiguous parts of the spectra (H1-2), but that (b) for human agent spectra the cognitive conflict maxima were more pronounced than for nonhuman agent spectra (H2-1) and the category boundary was prominently shifted towards the human endpoint (H2-2); see Figure 1. The data shows that that cognitive conflict processing peaks around the spectrum-specific category boundaries for all agent spectra independent of the agent’s human-likeness. However, both the extent of cognitive conflict processing and the location of the spectrum-specific

category boundary were affected by the specific categorization that needed to be made, that is: stimuli located at a nonhuman-human category boundary induced stronger cognitive conflict processing than stimuli located at a nonhuman-nonhuman category boundary, with no difference in the extent of cognitive conflict between non-alive-alive (i.e., robot-animal) and nonalive-nonalive (i.e., robot-stuffed animal) categorizations within the nonhuman spectra. These results suggest a quantitatively, though not qualitatively, different categorization process for spectra containing human endpoints.

The results are in line with previous studies that have linked the UV to categorical perception, with an increase of cognitive processing costs and decrease of affective ratings for categorically ambiguous stimuli (e.g., Cheetham et al., 2011; Mathur & Reichling, 2016; Weis & Wiese, 2017; Wiese et al., 2018). The universal observation of cognitive conflict processing for all tested spectra is in also in line with the inhibitory devaluation hypothesis (Ferrey et al., 2015), stating that phenomena related to “uncanny” stimuli, such as negative affective reactions, are not directly related to human-likeness per se, but reflect a more general form of stimulus devaluation that occurs when inhibition is triggered to resolve conflict between competing stimulus-related representations. Although no affective measures have been collected in the current study, the results show that common features of uncanny stimuli, such as an increase in cognitive conflict processing, are observable independent of considerations related to human-likeness, which provides support for the assumption that the observation of UV patterns might but due to a more general mechanism, such as cognitive inhibition.

*Eva Wiese (George Mason University)*

#### **Towards the Development of Semi-Supervised Rehabilitation Systems for the Home**

With the aging of the US population, there is an increasing need for effective and accessible rehabilitation services for debilitating illnesses and injury such as stroke [2]. Effective rehabilitation requires intense training and the ability to adapt the training program based on the progress of the patient and the judgment of the therapist [3]. Adaptive home based rehabilitation is emerging as a key avenue for improving health and wellness outcomes with the potential for reducing costs [1, 4]. While there are a number of challenges to scaling adaptive rehabilitation in the home (e.g. cost, motivation, adherence etc.), in this paper we describe our approach in aspiring to replicate the therapist presence and experience in the home. Here we describe our methodology in working within the nascent design space of evidence based adaptive healthcare. Evidence based adaptive healthcare has quantitative data at its core and can benefit greatly from advances in computation. Our focus is on designing adaptive home based rehabilitation systems for stroke survivors that can automatically track and assess complex human performance;

in our case adaptive rehabilitation training. Creating such systems requires a multidimensional approach integrating technical, medical, social, design, and human computer interaction knowledge.

We present our progress in co-designing an interactive system for upper extremity stroke rehabilitation in the home. Our interdisciplinary team of designers, engineers, computer scientists, doctors, patients, and therapists are iteratively designing, prototyping, and refining an experimental table top system for the home. While our goal is to ultimately create a robust system for lightly supervised use, we are first concerned with accomplishing key human-centered design activities that are essential for our development: i) analyzing and structuring performance for computational assessment; ii) collecting and assessing video data; iii) defining a movement performance rating rubric; and iv) rating the videos by experts. Findings from these activities will inform our algorithmic approach for the ultimate automated assessment of human performance in the home.

Analyzing and structuring performance for computational assessment: Working with the rehabilitation experts on our team, we developed a standardized set of exercises for upper extremity rehabilitation of stroke survivors. The set of exercises need to scale in complexity and map well to activities of daily living. To achieve the generalizability of a limited set of exercises mapping to many activities of daily living, they used an implicit segmentation of the exercises into types of segments that held true across all the exercises. For example, they looked at the characteristics of initiation, progression, or termination of movement (IPT) across all different types of exercises.

These exercises are realized using a custom-designed collection of modular objects that can be combined and configured in a variety of ways. We developed a prototype interactive system (see Fig. 2) for performing and capturing the exercises which comprises the objects, a custom mat, and a tablet interface delivering the exercise protocol (e.g. instruction videos indicating what objects to use and how to perform the exercise).

Collecting and assessing video data: We collected video recordings of nine stroke survivors attempting the 12 exercises while using the system in exploratory (non-therapeutic) study sessions at a rehabilitation hospital in the south east of the United States. Two women and seven men participated in the collection process, two with moderate impairment and seven with mild to moderate impairment. Each participant session was video captured through a two-camera setup (one sagittal for torso and shoulder capture, and one tabletop camera for capturing the wrist and fingers).

Defining a movement performance rating rubric: As there is a lack of consensus among therapists regarding the standardized, quantitative assessment of movement quality components and the influence of such components on overall functional ability [Levin], there is a need to create a consistent, quantifiable performance assessment rubric that is usable and appropriate for both human and computational agents. Our team prepared a simple web application for viewing and rating the collected video data from the nine participants. A team of four therapists spend two days viewing the data and collaborated to define a standardized movement segmentation approach for the exercises observed. Our team then built on this to create a more robust online system for rating movement performance based on this rubric.

Rating the videos by experts: Our online system supported 12 participating therapists in viewing and rating the video data (stored on encrypted USB drives shipped to the therapists) according to the

rubric developed by our therapist team (see Fig 1.). Figure 3 depicts the movement segmented interface, where therapists can rate and rationalize their assessments.

We are currently assessing the inter-rater reliability of this work and analyzing the findings from an online survey review and a two-day workshop with the assessment team. Findings from these results can help improve the assessment rubric, the rating interface, and the overall performance of the therapists in understanding the nuances of their own therapeutic approach and their ability to make standardized evidence-based decisions.

*Aisling Kelliher (Virginia Tech University)*

### **Technology Made Us (Motion) Sick; Autonomous Cars Will Make Us Vomit: A Review of 40 Years of Research**

Man was built self-propulsion. It was technology that introduced the option of traveling by carts, ships, cars and airplanes. Thus, it was technology that produced seasickness, car sickness, and air sickness. Researchers, including many psychologists, have spent years seeking solutions for this ancient ailment that persists in modern times. Despite their efforts, experts predict that the introduction of autonomous vehicles will dramatically increase the prevalence of motion sickness.

At the Naval Motion Sickness and Human Performance Laboratory, we have been exploring the psychological and physiological aspects of motion sickness for almost 40 years, and have published numerous articles on this subject. In this lecture, I will review the knowledge accumulated in this field and focus on some technological and psychological solutions we have pursued over the years

The extent of the problem. Findings from our laboratory and from many other studies show that nearly everyone can suffer from motion sickness. Though significantly more prevalent when traveling by sea, airsickness and carsickness are very common as well.

The sopite syndrome. A less-researched phenomenon is the lethargy, apathy and sometimes even depression that emerge during exposure to uncontrollable motion. We studied this phenomenon based on the learned helplessness paradigm and were able to show significant cognitive, emotional and performance deficits.

Why are drivers rarely motion sick? We were the first to study the role of controllability in motion sickness. We compared groups that differed in their controllability of head movements, vision, activity, and predictability. Literature has suggested these factors as explanations for the driver's immunity to motion sickness. Subjects who were given more control reported significantly fewer motion sickness symptoms and a smaller decrease in their well-being, compared to the yoked subject who lacked control.

Based on these findings, we predict that the prevalence of carsickness in autonomous cars will be significantly higher.

Autonomous cars—More sickness by far. Autonomous cars will soon be an integral part of our lives, and it is clear that they will present new challenges and disturbances for their passengers.

This change will likely raise issues related to trust. At least at the beginning, people may tend to be somewhat apprehensive about being moved by a computer.

Lack of control and inability to predict the direction of motion will be the main factors that contribute to enhancing motion sickness in autonomous cars. Moreover, passengers will likely use the time for reading or watching videos, which will certainly cause a significant increase in carsickness as well.

Technological solutions. The lecture will review various ergonomic and human factors that can alleviate motion sickness. One example presented will be our studies on the artificial horizon.

It is well-known that motion sickness is caused by visual-vestibular conflict. In modern navy ships, for example, most of the crew is stationed beneath the deck and suffer more than those who are above-deck and can see the horizon. We studied the possibility of reducing this conflict using an artificial horizon. We used a tilting room to investigate the potential beneficial effect of using a projected artificial horizon as a means to prevent seasickness. Indeed, there were fewer motion sickness symptoms when artificial horizons were used. The visual reference prevented the decline in performance found under the “closed cabin” condition. The results of this study suggest that a projected horizon might alleviate motion sickness.

Psychological solution. There is ample of evidence that psychological factors can reduce motion sickness. Some of our field studies showed the possibility of using cognitive behavioral interventions to cope with motion sickness.

We examined the general repertoire of self-control behaviors and their ability to cope with seasickness. We found that subjects with high self-control reported using self-control methods to cope with seasickness more extensively than did subjects with low self-control.

In his book on motion sickness, published in 2019, Dobie summarizes his career in the British and the Canadian air force as well as in the US Navy. He describes CBT as the main approach for coping with motion sickness. It is predicted that as a significant percentage of people will likely suffer from car sickness, CBT will be very much needed in the era of autonomous cars.

Biofeedback. A series of NASA studies showed that biofeedback training can prevent motion sickness in space. We followed up on this line of research with navy personnel and found that biofeedback can reduce the helplessness associated with aversive motion. Other studies from our lab have shown the importance of R-R variability in motion sickness, which suggests that Heart Rate Variability biofeedback can reduce motion sickness even further.

*Arnon Rolnick (Rolnick's Clinic)*

### **Collision or Avoidance: Should Auditory Warnings Be Toward Danger or Toward Safety?**

Background: Auditory warning systems help drivers notice potential hazards and avoid upcoming threats safely (Spence & Ho, 2008; Mohebbi, Gray, & Tan, 2009). However, the directionality of warning signals is often underutilized in typical warning systems (Beattie, Baillie, Halvey, & McCall, 2014; Sabic & Chen, 2017). For example, to alert the driver about a pedestrian who is crossing the road, a warning can signal the direction of a potential collision (i.e., the location of the pedestrian) so that the driver can easily locate the pedestrian (collision direction; Wang, Pick, Proctor, & Ye, 2007; Straughn et al., 2009), or it can signal the direction toward which the vehicle should be maneuvered in order to avoid the pedestrian (avoidance direction; Wang, Proctor, & Pick, 2003). A recent study by Sabic and Chen found that the presentation of a tone in the collision direction resulted in more efficient responses of the drivers than the presentation of a tone in the avoidance direction.

Research problem: To follow up Sabic and Chen's (2017) study, we examined whether the effectiveness of warning tones in the collision and avoidance directions depended on how close the driver

was to a potential collision. We tested this by varying the interval between the presentation of a warning tone and the time at which the vehicle would hit the pedestrian (times-to-collision, or TTCs).

Procedure: Forty-five participants (39 female; age  $M = 19.79$ ,  $SD = 2.67$ ) were tested in a simulated semi-automated driving task. Participants avoided pedestrians walking in the road by turning the steering wheel to the left or right. A warning tone was presented in either the avoidance or collision direction at the same time as the appearance of a pedestrian. The warning direction was a between-subjects manipulation to avoid confusion between the different warning methods. One key difference from Sabic and Chen's (2017) study was that the pedestrian appeared in the middle of the road and walked to either side, rather than appeared from either side of the road and walked to the other side. This manipulation was to control the possible confounding factor introduced by the starting location of pedestrians in Sabic and Chen's study. TTC varied between 1.5 and 3.5 seconds with 0.5-second interval. Reaction time (RT) and accuracy were recorded to measure the effectiveness of the warnings.

Analyses: We ran separate 2 (warning direction: accident, collision; between-subjects)  $\times$  5 (TTC: 1.5, 2.0, 2.5, 3.0, 3.5 seconds; within-subjects) mixed Analyses of Variance (ANOVAs) on RT and accuracy, with the Greenhouse-Geisser correction when the sphericity assumption was violated.

Results and discussion: For RT, the main effect of TTC was significant,  $F(1.78, 76.62) = 128.59$ ,  $p < .001$ ,  $\eta^2 = .749$ , suggesting that RT increased as TTC increased ( $M_s = 722.74, 787, 869, 935$ , and  $998$  ms, for the 1.5-s to 3.5-s TTCs, respectively). Pairwise comparisons with Bonferroni correction showed that RTs differed across all TTC levels,  $p_s < .001$ . The main effect of direction was not significant,  $F(1, 43) = 2.53$ ,  $p = .119$ ,  $\eta^2 = .066$ , ( $M_s = 9029$  vs.  $816$  ms for avoidance vs. collision directions). There was no significant interaction between TTC and warning direction,  $F < 1$ . For accuracy, there was no significant effect,  $F_s < 1$ . Accuracy was high in both conditions ( $M_s = 99.0\%$  vs  $99.1\%$  for collision vs. avoidance), demonstrating a ceiling effect.

The 94.77-ms difference between the two warning directions in the current experiment was not statistically significant. The discrepancy from the 202.92-ms difference found in Sabic and Chen's study is likely due to the change in the pedestrian's starting position. When the pedestrian appeared on one side of the road and started walking towards the middle, as in Sabic and Chen's study, the collision warning was on the same side as the pedestrian whereas the avoidance warning was on the opposite side. This condition provided stimulus-stimulus compatibility between the warning and the pedestrian locations (Kornblum, 1992), and the warning signal cued the location of the pedestrian. The spatial compatibility between the warning and the pedestrian facilitated responding. When the pedestrian appeared in the middle and started walking towards one side, as in the present study, neither warning location was compatible with the pedestrian's location, thus leading to a non-significant difference between the two warning directions. The results of the two studies together imply that the warning is more effective when it signals the current location of the pedestrian rather than a predicted position at which a collision might occur.

Conclusions and practical implications: The shorter TTC conditions resulted in faster RTs, showing that the trend of more impending TTCs increases RTs. The present data showed a non-significant advantage of the collision direction warning. Combined with previous findings in Sabic and Chen (2017), we recommend that the warning should signal the location of the pedestrian.

*Scott Mishler (Old Dominion University)*



## Older Adults and Technology

### PAPER SESSION

#### The Oldest Olds' Perceptions, Attitudes, and Acceptance of Future Transportation Technologies

The world's population is experiencing unprecedented longevity. In the United States alone, the share of the population aged 65 and older increased from 12.4% to 15.6% between 2000 and 2017 (PRB, n.d.). Meanwhile, the 85 and older age segment is projected to more than triple from 6 million people today to almost 20 million people by 2060 (Mather, Jacobsen, & Pollard, 2015), making it fastest-growing age segment in the United States (Ortman, Velkoff, & Hogan, 2014). The impact of this demographic shift on transportation—and vice versa—will be monumental. Transportation is essential to the ability of older adults to live independently, maintain mobility, access services and institutions, avoid social isolation, ease burdens on family or society, and have spontaneity in daily life. The proliferation of older road users will lead to new challenges in terms of safety, traffic, and sustainability (Chan, 2017). There have been considerable advancements in automated driving technologies over the course of the last decade. Given the growing number of older adults, the future of transportation technologies promises an aging population greater independence and mobility than ever before (NSTC, 2019). Understanding the general implications of population aging has received significant attention, but little research has been done with the “oldest old,” those age 85 and older. Similarly, despite enthusiasm and speculation about how emerging automotive technologies will affect the future of transportation, there is less known on the factors underlying consumer perception and acceptance of these new technologies across age demographics. Nees (2016) found age was slightly negatively correlated with self-driving car acceptance. In 2018, Lee et al. reported fewer older individuals (65+) chose full self-driving as the maximum level of automation they would be comfortable with compared to younger groups. In the same study, however, the majority of those over the age of 55 indicated high comfort with driver assist features (Lee et al., 2018). While this literature offers a foundation for future research, it does little to uncover the underlying factors that contribute to comfort with and acceptance of automated driving technologies among specific consumer groups that stand to benefit the most from improvements to future vehicle technologies.

Technology adoption and acceptance has historically been understood through the lens of theories such as the Diffusion of Innovations Theory and Technology Acceptance Model. While these frameworks often highlight the complexity of automobile technology adoption as a systems issue, they lack the integration of individual psychological factors and the concept of time in understanding human behavior. Technology adoption is inherently impacted by our psychology and may change as we age (Parker & Manstead, 1996; Porter, 2011).

Feelings impact how people make decisions about about transportation, risk perception, benefit perception, and trust in technology. Previous research has demonstrated the importance of affect when people make judgments (e.g., Zajonc, 1980), and in particular the role of feelings in the evaluation of risk-benefit tradeoffs (Peters,

Västfjäll, Gärling, & Slovic, 2006; Slovic & Peters, 2006; Tompkins, Bjälkebring, & Peters, 2018). Affect becomes more prominent under conditions characterized by higher levels of uncertainty, less familiarity, and greater complexity (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Faraji-Rad & Pham, 2016; Srull, 1984). Further, negative affect has been shown to have more influence shaping judgments than positive affect (Baumeister et al., 2001). Raue et al. (2019) found that feelings related to traditional driving affected individual perception and acceptance of self-driving cars. Taken all together, these results suggest that feelings may be significant factors in the decision-making processes among the oldest old when it comes to technology adoption behind the wheel.

This paper will present findings from mixed methods research with the Lifestyle Leaders, a bimonthly panel of adults ages 85 and older at the MIT AgeLab. The purpose of the Lifestyle Leaders Panel is to explore aging issues with those who are on the “leading edge” of aging. The paper will analyze data collected from questionnaires and focus groups with the Lifestyle Leaders about transportation technology adoption and acceptance, especially trust and use of in-vehicle technologies and self-driving vehicles.

Preliminary analyses suggest the oldest old associate convenience, freedom, and enjoyment with driving. These feelings toward traditional driving influenced greater levels of comfort with self-driving cars. Higher levels of physical discomfort when driving were positively correlated with comfort with self-driving cars. Qualitative data indicate older drivers view in-vehicle technology and self-driving vehicles as a way to stay driving, remain independent, and reduce risk. This paper will report on these and other factors that impact in-vehicle technology and self-driving vehicle adoption among a panel of adults aged 85 and older, resulting in a better understanding of the impact of emerging automotive technologies on shifting age demographics and resultant changes to driving behavior.

*John Rudnik (Massachusetts Institute of Technology AgeLab)*

#### Optimizing Aging-relevant Technologies: A Person-centered Approach

Statement of the Problem: Innovation informed by behavioral and social science research is necessary to successfully exploit technological advances and their potential to contribute to optimal aging. A critical component of designing technologies for aging adults is to understand and respond to the goals, needs, and preferences of heterogeneous aging adult cohorts (e.g., Peek et al., 2015). Procedures: To investigate the correspondence of optimal aging and technology perceptions, we surveyed 604 participants enrolled in the RAND American Life Panel, a nationally representative, probability-based panel of 6,000 adults. Participants ranged in age from 50-90+, 51.5% were female, 44.7% were employed, 59.6% were married, and 50% reported educational attainment of bachelor's level or higher. Overall, 8% of the sample identified as Hispanic, with 15% of individuals also identifying as Black, Asian, or Asian Indian or Alaskan Native. Analyses and Results: The first aim of this paper was to characterize respondents' optimal aging goals and concerns. The results indicate that the vast majority of respondents (85%) expressed interest in how to age well. The optimal aging goals most frequently endorsed included the ability to take care of oneself and maintenance of good health until close to the end of life as well as being able to act according to personal standards and values (>92%). Regarding future needs, 75% of the participants expressed specific

concerns about their future functional ability. The second aim of the study was to characterize participants' reported knowledge and openness to aging-related use of specific technologies. Familiarity with a diverse set of technologies that could optimize aging varied (e.g., 33% augmented reality to 95% voice activated tech), as did respondents' inclination to use technology to age well (46-69%). To investigate further, we employed regression analyses to identify significant predictors of technology use for optimal aging including sex, self-reported health, concerns about future functioning, and technology optimism and innovativeness. Discussion: Discrepancy was evident between familiarity and anticipated use, perhaps reflecting skepticism and misconceptions about more familiar technology (e.g., smartphone, voice activation) and greater optimism for less known technology (e.g., home sensors, telehealth). Discussion focuses the value of maximizing the match between adults' everyday goals and concerns with specific technology platforms.

*Jennifer Margrett (Iowa State University)*

### **Interacting with a Robot Partner: Age-Related Differences in Communicative Perspective Taking**

The expanding aging population and the associated demand on healthcare and economic resources has led to an increase in research on technologies to support healthy aging. Many of these technologies involve intelligent agents equipped with spoken language interfaces (e.g., SmartHomes, social robots). However, the design focus has mainly been on appearance and voice recognition/speech clarity rather than cognitive processes that underlie successful communication. In the current study, we examine to what extent younger and older adults engage in communicative perspective-taking while interacting with a social robot. We use gaze measures to track the fine-grained time course of interpretation in human listeners, as well as listeners' implicit hypotheses about the linguistic intentions of the robot. Our specific focus is the extent to which human listeners spontaneously ascribe human attributes to the robot and use this belief to guide interpretation as spoken language unfolds in real time.

Seventy-two adults (36 younger, Mean=19.97; 36 older adults, Mean=72.61 years) participated in study. Participants followed instructions from a social robot (Furhat Robotics, Sweden) to click on objects on a screen. On critical trials, the name of the intended target ("blacksmith") overlapped with a color term ("black"). One question was whether listeners could remember a previously stated limitation of the robot (that it cannot see color) and use this knowledge to guide real-time language processing. A second question was the extent to which listeners assumed the robot would follow the kinds of pragmatic rules obeyed by humans in language use (using an adjective like "black" in a description primarily when there is a genuine need to differentiate a referent from another same-category alternative). Across conditions, we varied whether the target bore the color property in the overlapping adjective (whether the blacksmith had a brown or black tunic). An increased tendency to rapidly shift gaze to the target in the latter case would reflect listeners' difficulty in reducing attention to color attributes. We also varied whether a competitor object in the scene (black flag) was accompanied by another same-category object (gold flag). When this contrasting same-category object is present, this should boost the likelihood of misinterpreting the beginning of "blacksmith" as an adjective (e.g., the beginning of "black flag"), but only if listeners spontaneously assume

the robot follows human linguistic-pragmatic conventions (such that the competitor would instead be called "the flag" in the absence of a contrasting item in the same category). Finally, we recruited two different age groups (younger and older adults) to explore age-related changes in real-time language processing with the robot partner. Of particular interest is the fact that older adults are often described as having reduced working memory (Wingfield et al., 1988) and inhibitory control ability (Hasher & Zacks, 1988), which may impact their ability to "bear in mind" the limitations of the robot partner.

There were 12 critical trials, interleaved with 24 filler trials. We had four conditions, created by crossing the target color manipulation (e.g., whether the blacksmith's tunic was brown or black) with the competitor contrast manipulation (whether the black flag was accompanied by another flag of a contrasting color, Figure 1). Participants were seated in front of a 40" monitor, with the robot (Figure 2) placed to the right side. Participants' point-of-fixation was monitored using an EyeLink 1000 eye tracker. Before the experiment began, the robot engaged in some conversational small talk, and the eye tracker was then calibrated.

We conducted growth-curve analyses in R to examine the extent to which listeners considered the competitor object (black flag) relative to the target object (blacksmith) as the target word unfolded in time. The analysis window was 1000 ms in length starting from the onset of the noun. The results revealed that both younger and older adults were more likely to consider the competitor when the target (blacksmith) bore the color of the competing adjective term (black tunic), reflecting a fairly automatic attentional re-orienting due to the semantics of the color term (see Huettig et al., 2011 for review). Further, an effect of the competitor contrast manipulation was evident in older adults, such that they were more likely to make fixations to the competitor (black flag) when it was accompanied with a color contrast object (gold flag) than when it was accompanied with an unrelated object (strawberry). This specific result indicates that older adults have more difficulty than younger adults when bearing in mind the robot's limitations (and instead implicitly assumed it might use a color term to differentiate same-category objects). This is consistent with the notion of age-related decline in working memory and inhibitory control.

Together, the findings highlight underlying mental processes involved in human-robot communication, and how these differ across the adult lifespan. This research helps provide the scientific grounding needed to inform the design of assistive technology and other adaptive technological communication aids for the aging population.

*Raheleh Saryzadi (University of Toronto)*

### **Autonomy Through the Autonomous: Increasing Older Adults' Self-Sufficiency with Self-Driving Cars**

The car is a ubiquitous feature of American life; the majority of American adults—including adults ages 65 and older—drive a car most days. With increasingly advanced and available autonomous technology, fully autonomous vehicles are likely to reach the consumer market in the near future. Regulators, manufacturers, and advocacy organizations predict that increased autonomy for vehicles will also increase autonomy for users of these vehicles, particularly by allowing older adults to maintain transportation independence.

In this United States, which has a geographically-disparate and rapidly-aging population, access to transportation is crucial in allowing older adults to live independently. However, regulatory paradigms

have not yet fully considered the substantial and unique challenges that will face both older adults operating autonomous vehicles and the legal system governing them.

This paper provides a brief background on autonomous vehicles, discussing both scientific and societal challenges to their widespread introduction. It outlines features and limitations of autonomous vehicle technology, noting areas of overlap with the needs of older adult drivers. This paper then considers older adults' transportation needs, addressing legal challenges related to capacity and guardianship, physical and cognitive limitations related to aging, and regulatory challenges related to driver licensure and operation of both traditional and autonomous vehicles.

Next, this paper analyzes the implications of regulatory structures that would focus on autonomous vehicles and older adults. It then considers the positions and goals of stakeholder groups, including the unique perspectives of older adults as potential consumers and users of autonomous vehicles. Finally, the paper identifies paradigms for framing funding and liability issues related to autonomous vehicles. This paper concludes by considering under what circumstances autonomous vehicles will be most likely to effectively increase independence for older adults.

*Mary Nikityn (Penn State Dickinson Law)*

## Independence HI



### Games for Growth: Innovative Approaches to Applying Digital Games in Therapy

#### PANEL

Video games are immensely popular. More than half of the people in the U.S. play video games and that number rises to 90% when considering younger demographics. But games aren't just for kids. The average age of a gamer is 34 and gamers over 50 are the fastest-growing consumer demographic. Sixty-seven percent of parents report playing video games with their kids at least once a week. Sixty-four percent of US households own a gaming device and 60% report playing games daily. 45% of gamers in the US are women. For the last several years, the video game industry has generated more revenue than the television and film industries combined. In short, games are a big deal and there is a very high likelihood that any client walking through the door at plays some kind of video game on a regular basis.

The sheer volume of people playing video games, both in the US and globally, means that understanding games and games culture is a matter of competence for today's mental health professionals. Knowing what game your client enjoys and why they enjoy it offers a treasure trove of potential insight and being games-literate is an "instant rapport button" for clients who value their gameplay experience. Ignoring or avoiding the role games play in a client's life as well as in society is no longer an ethical option.

This interactive panel is comprised of a unique set of experts who bring qualifications in both games and mental health to the table. Each presenter will provide a brief snapshot of their intersectional work and how being culturally and technologically mindful of games has the power to improve therapeutic alliance, treatment adherence, rapport-building, goal-setting, and therapeutic outcomes. This panel is not about how games are a panacea; instead presenters

will share, from their own research and experience, how games are a tool for growth. Topics covered include a wide range of technological implementations, from VR to pen and paper.

In addition to the diversity of tech discussed, this presentation also offers a diversity of perspectives. Presenters include social psychologist and Yale alum Susan Rivers, Ph.D., licensed play therapist and psychologist Jessica Stone, Ph.D., Eleanor Roosevelt Human Rights Award winner and published author Janina Scarlet, Ph.D., clinical director of the non-profit mental health and games advocacy charity Take This Raffael Boccamazzo, Psy.D., and psychologist and game designer Kelli Dunlap, Psy.D. whose work on mental health and games has been featured by leading organizations such as UNESCO and the New York Times.

(All stats retrieved from the 2018 ESRB Essential Facts: [http://www.theesa.com/wp-content/uploads/2018/05/EF2018\\_FINAL.pdf](http://www.theesa.com/wp-content/uploads/2018/05/EF2018_FINAL.pdf)).

*Chair: Kelli Dunlap (iThrive Games)*

#### Superhero Therapy

This talk will discuss how to incorporate a role-playing game, Therapy Quest, into evidence-based therapy, such as cognitive behavioral therapy (CBT) and acceptance and commitment therapy (ACT) in treating a variety of mental health disorders, such as PTSD, anxiety, and depression. Given that many young people struggle to label and understand their emotional experiences, the Therapy Quest approach will demonstrate how therapists can help clients to become an active participant in an adventure to learn about their mental health and learn mental health coping tools. Due to the fact that this is a novel approach, the direct empirical evidence related to its effectiveness is limited right now. However, the Therapy Quest approach is grounded in the neuroscience of play behavior, as well as the empirical research on role playing games, and its adaptation in evidence-based therapies. In addition, this talk will provide an explanation as to how these role-playing games can stimulate prosocial behaviors, which are important in development and establishment of healthy social support.

*Janina Scarlet (Center for Stress and Anxiety Management)*

#### Integrating Technology into Modern Therapy

Clinician decisions to utilize digital tools in therapy must be made with great care and forethought. The appropriate therapeutic use of digital tools in sessions can enhance the rapport, assessment, and intervention portions of treatment. This presentation will focus on the use of tablets, the Nintendo Switch, and virtual reality in session. Research, brief case examples, and information on a few specific programs will be introduced.

*Jessica Stone (Virtual Sandtray)*

#### Museum of Me

There are too few appealing, developmentally appropriate methods for engaging teens in their own social and emotional learning. Commercial video games offer rich, mediated, interactive narrative experiences that can be integrated into the classroom to support teen students' core academic knowledge as well as their social and emotional development. This paper describes how and why educators are using the award-winning video game and digital "museum,"



What Remains of Edith Finch, to engage high school students in building both literacy skills and social and emotional skills. Players of What Remains of Edith Finch interact with the narrative told by an 18-year-old girl while exploring her childhood home, which has become a museum preserving the identities of family members who have mostly succumbed to tragic fates. This talk will demonstrate how this digital museum can be used with teens to explore how they become aware of and represent their identities publicly and privately. Lessons that include gameplay, artifact creation, and rich discussions about identity of self and others will be shared.

*Susan Rivers (iThrive Games)*

#### Cheat Code:

#### Integrating Digital Games in Therapy without Tech

Over half the US population plays video games. Although the number varies based on demographic, it is sufficient to say that video games are popular and here to stay. For many, games are a way to cope with life's stressors; the majority of modern games are highly social and provide ample opportunity for challenge and mastery as well as relaxation and release. While games are not a replacement for therapy, they can be leveraged to help a client reach their treatment goals the same way a therapist might utilize a client's love for a specific book or sport. The trouble is, there are thousands of games and it's impossible for even the most games-savvy clinician to keep up. Factor in additional complications that frequently arise when integrating games into therapy (i.e. tech issues, outdated software, time constraints, etc) and it can seem like an insurmountable problem. This talk will briefly highlight the above challenges and then provide an overview of the process undertaken to confront and resolve them. Specifically, the talk will layout the year-long development process of "Clinician Game Guides" - a collection of one-sheet guides that are designed to provide clinicians with the cliffnotes version of popular games such as Fortnite and Overwatch. Each guide was created by a mental health expert, reviewed by a licensed psychologist, and features a game summary, therapeutic themes, discussion questions, and a case study. Attendees will walk away from the talk with a deeper understanding of the multidisciplinary collaborative design process that made the guides possible, access to the guides themselves, and foundational skills on how to integrate AAA games into a therapeutic setting without having to be gamers themselves.

*Kelli Dunlap (iThrive Games)*

#### The Need for Geek and Gamer Cultural Competency

Video games, superhero movies, online streaming, and other pursuits commonly viewed as "geeky" have experienced an astronomical surge in popularity over the last 20 years. Video games alone account for over \$30 billion in revenue. Despite the increase in popularity, how many clinicians know about the experiential differences when clients talk about being a YouTube content creator, casual roleplaying game fan, or a collegiate eSports player, among other pursuits? Do these all fall under the monolithic construct of "screen time", or are there more nuanced differences? If we are not able to understand the experiences of our clients, we cannot serve their needs. This talk will provide high-level overviews over some of the experiences in which our clients engage, and why thinking about being a "geek" in terms of cultural competency behooves us all.

*Raffael Boccamazzo (Take This)*

#### Lafayette Park

### Machine Learning For Analyzing Behavior

#### PAPER SESSION

#### Using Machine Learning to (Re-) Describe, and Predict Mental Ill Health

Owing to the increasing prevalence, and changing nature of mental ill health, there is an emerging urgency to improve our understanding of mental illness and develop an ability to predict the risks, and course of psychopathology. This shall inform clinical decision-making over treatment, preventive interventions, and support services. There is also increasing intellectual dissatisfaction with the traditional conceptualization of mental illness, among both researchers and clinicians, and across the globe. The study intends to present a novel approach to using Machine Learning to achieve a new understanding of mental ill health. Advancing on this knowledge, the next step builds up a pipeline to test how accurately individuals can be classified into one of the clusters, and predict the risk of psychopathology for each individual.

Methodologically, the first part of the study used Clustering algorithms as an exploratory text-analysis of patient's first-person narratives (N= 222) about their lived illness experience to discover latent structure (or themes) and accordingly organize illness narratives into meaningfully similar groups. We employed a frequency search, followed by text-based cluster-analyses, using the complete linkage (farthest neighbor) hierarchical clustering algorithm. Combined, the procedure has lent insight into novel patterns, and groupings of the mental ill health conditions that could subsequently lead to a preliminary re-conceptualization of existing psychiatric categories.

We intend to advance this preliminary study by increasing the volume of data, and converting the corpus to TD-IDF and then applying clustering algorithms to cluster the documents, using the metrics of Cosine distance as a part of unsupervised machine learning pipeline implemented in NLTK and Scikit-Learn. We also aim to improve the performance of the model using the addition of contextual feature extraction.

In the second part of the study, we aim to put the discovered clusters under test for external validity by estimating the accuracy with which the algorithm is being able to classify individuals into one of the discovered clusters (each of which has its own description). To achieve that, we intend to design a multiclass classification algorithm. Once trained, we will be able to pass a new patient narrative (i.e. a document) to the predictive algorithm and get a diagnosis (as output) for the narrative. Finally, we intend to run a cross-validation process to select the best performing model and make predictions. Additionally, metrics such as precision, recall, accuracy, and F1 scores will be used to guide our selection of algorithms.

The preliminary results for the first study (can be found on the attached PDF with the submission) were based on the frequencies, and occurrences of the words in the text (corpus) representing conditions (e.g. symptoms), or events (e.g. rape). They were analyzed and presented in the form of a word cloud and dendrogram.

The Word Cloud depicts up to the 228 most frequently occurring stemmed words (up to 0.5% word frequency) in varying font sizes, where frequently occurring words are in larger fonts. For example, the stemmed word "depressive" has occurred the most number of times

in the corpus. All the similar words like 'depression', and 'depressed' are included under the stemmed word "depressive."

Likewise, dendrogram represents words that co-occur and are clustered together. For example, words like 'drinking' and 'alcohol' co-occurred and hence been clustered together. New insights can be formed from such clusters. For example, patients reporting eating disorders has also reported alcohol drinking more often forming a cluster that can be termed as "Conditions related to Consumption." This is interesting because historically, diagnostic systems have treated 'eating disorders' as different from alcoholism, but patients in this corpus seem to have reported experiences of these two more often.

Finally, we have compared the similarities, and dissimilarities between each document (illness-narrative) using the Pearson correlation coefficient and are presented in the form of a Scatterplot. Cases of patients that have a higher degree of similarity based on the occurrence and frequency of words were clustered together. While those who have a lower degree of similarity based on the occurrence and frequency of words were displayed further apart. It is noteworthy to mention that we discovered that the patients diagnosed with Major Depressive Disorder, Obsessive Compulsive Disorder, Post-Traumatic Stress Disorder, and so on, did not form clusters on their own but were found to be spread far and wide in the scatterplot. This indicates a high degree of dissimilarities between the narratives of patients diagnosed with the same diagnostic label.

To sum up, the first study aims to contribute to the current debate on the definition of (and potentially redefining) mental ill health. The second study promises the implication of these two studies in the form of a potential clinical diagnostic support tool to help clinicians make informed clinical decisions.

*Chandril Ghosh (Queen's University Belfast)*

### **Evaluating Team Communication via Audio Segmentation and Interaction Identification**

Measuring characteristics of teams is a challenge for at least two reasons. The first difficulty is that attributes such as team trust and team communication are difficult to quantify. The second reason is that overall task performance results from a combination of the team's attributes and the team members' skills with the task at hand, so that comparing teams based on performance of a single task makes little sense. The present research addresses both difficulties for the attribute of team communication.

When analyzing a team, it is important to distinguish between team skills and task skills. Team skills are the abilities of team members to function effectively as a team regardless of any specific task. Examples of team skills are shared cognition and coordination (Prichard, Stratford, & Bizo, 2006). In contrast, task skills are the skills that members of the team possess specific to the task at hand. For example, it would be difficult to use one soccer game (a task performance) to compare the team skills of an amateur soccer team that has played together for years with those of a newly-formed team of professional players. Instead, evaluation methods that focus specifically on assessing team skills directly are more appropriate.

This research effort attempts to quantify the team skill "team communication." A common solution is to define abstract behavioral constructs in terms of observable events known as behavioral markers, e.g., sharing information or owning up to a mistake. Sottilare et al. (2017) identified behavioral markers for several team-related constructs. However, the process of detecting instances of markers in

practice is a deceptively difficult task unto itself and is still an active area of investigation (Bonner et al., 2016; Flin & Martin, 2001). For example, using an automated approach to measure communication by counting team member utterances would not take interactions between team members into account and offer a deficient evaluation. Such a measure would also not differentiate whether a frequent speaker is dominating the conversation or is serving as a helpful facilitator.

This research offers a novel approach to evaluate interactions between team members directly by analyzing patterns of speech timing. An interaction event was defined as a sequence of communications regarding a single idea, including statements and acknowledgments. Whereas some speech processing systems make use of Hidden Markov Models (HMMs) for audio data analysis, this was not feasible for the stated application due to an HMM's requirement for extensive tagged data for training. Instead, the system used unsupervised classification methods. First, short-term features of an audio recording of two speakers were classified via an SVM model to segment the audio into individual utterances (Zahid, Hussain, Rashid, Yousaf, & Habib, 2015). Next, utterances were clustered by speaker using k-means clustering on a dynamically-optimized subspace of audio features (Giannakopoulos & Petridis, 2012). Finally, to identify sequences of utterances as interactions between team members, utterances were represented as nodes in a multipartite graph that could then be solved as a modified instance of the assignment problem from combinatorial optimization (Jiang, Ren, & Hu, 2008; Kuhn, 1955). Each of the resulting "assignments," e.g., a grouping of two or more related utterances, was considered a distinct interaction event between team members that could then be evaluated using team behavioral markers or other pre-established criteria.

The implementation was evaluated using audio recordings from a two-person virtual team task. Speaker identities, utterance timings, and interaction events were manually tagged with sufficient inter-rater reliability for comparison with automated audio analysis results. Evaluation of the automated interaction assessment system are in progress, with final results to be shown during the presentation. Initial results are promising.

This research represents a novel approach for automated evaluation of the team skill "team communication." Because the described method analyzes team communication based on interactions rather than inferring it from task performance, the approach is applicable to any task domain that requires communication and is not biased by individual team members' task abilities.

*Alec Ostrander (Iowa State University)*

---

## **Independence E**

### **Rethinking Anthropomorphism: The Antecedents, Unexpected Consequences, and Potential Remedy for Perceiving Machines as Humanlike**

#### **SYMPOSIUM**

Anthropomorphism—the attribution of "humanlike characteristics, motivations, intentions, or emotions" to nonhuman agents (Epley,

Waytz, & Cacioppo, 2007)—has ancient roots in human psychology, yet the rapid development of Artificial Intelligence (AI) and robotics technology is presenting this human tendency with new significance and unprecedented challenges (Broadbent, 2017). As machines and algorithms are often intentionally designed to look and act humanlike, their increasing prevalence in homes, market places, and public spaces raise pressing questions on the causes and consequences of anthropomorphism for psychologists, designers, marketers, and policy-makers.

What leads people to perceive machines and algorithms as humanlike? How does an increasingly blurred line between humans and machines influence people's interactions with machines as well as their perception of humanness? How do we overcome the potential pitfalls in people's natural tendency to anthropomorphize? In three empirical projects and one theoretical paper, this symposium invites the audience to rethink the age-old human tendency of anthropomorphism in light of the increasing prevalence of robots, chatbots, virtual assistants, and smart home devices in modern times.

The first paper by Zhao, Phillips, and Malle ("Beyond Anthropomorphism: Differentiated Inferences About Robot Mind from Appearance") explores how robots' humanlike appearance systemically influences people's beliefs about the robots' humanlike mental capacities. By creating and analyzing the Anthropomorphic roBOT (ABOT) Database—an image collection of over 250 anthropomorphic robots—the authors find that, in contrast to the common idea that "looking human" is a unidimensional concept (i.e., on a spectrum from "machine-like" to "human-like"), three distinct dimensions underlie robots' humanlike appearance (Body-Manipulation, Surface Look, Facial Features). Furthermore, the perception of humanlike mental capacities also consists of three distinct dimensions (Affective capacities, Social-Moral capacities, and Reality Interaction capacities). Intriguingly, specific dimensions of a robot's physical appearance trigger people's attributions about specific dimensions of its mental capacities. People's beliefs and impressions of machines' mental capacities can have important consequences, ranging from misplaced trust and attachment in the case of overestimations, to discomfort and shock in the case of underestimation. Therefore, this research challenges the common notion that people simply perceive a more human-looking agent to have a more humanlike mind and has important implications for designing social robots that elicit only those expectations that are consistent with their actual capacities.

Next, the paper by Bluvstein, Zhao, Barasch, and Schroeder ("Hello! How May I Helo You? How Written Errors Humanize a Chatbot") focuses on the widely used virtual agents in consumer service. Over three studies, the authors investigate how seemingly incidental features in online conversations can lead people to perceive a chatbot as humanlike and highlight its interpersonal consequences. Their research suggests that observing a conversation agent making and subsequently correcting conversational "mistakes"—like a typographical error—can lead people to attribute more humanness to the chatbot, perceive it to be warmer, and consequently, share more personal information with it. These findings provide insight into when people might share personal information, with potential implications for consumer online privacy.

The third paper by Kim and McGill ("Two Minds Meeting in the Middle: How Humanlike Technologies Lead to Dehumanization") explores an ironic relationship between anthropomorphism and dehumanization: How exposures to and experiences with humanlike technologies change people's perceptions about humanness. Over

four empirical studies, the authors found that, as people perceive technological products as more humanlike (i.e., having more mind), they perceive and treat actual people as being more object-like, going so far as to advocate dehumanizing behaviors toward others. As new technologies are increasingly designed and portrayed as humanlike for market appeal and ease of use, this research raises important concerns regarding how such technologies and narratives may introduce unintended social consequences.

In the concluding paper ("Object-Oriented Anthropomorphism as a Mechanism for Understanding AI"), Hoffman and Novak offer a potential remedy to the unintended consequences and potential dangers of anthropomorphism. To understand the experience of smart objects, they propose object-oriented anthropomorphism as an alternative metaphor to human-centric anthropomorphism. Object-oriented metaphor allows consumers to translate data about an object's behavior into a working model about how objects experience environments. A range of positive outcomes of object-oriented metaphors are proposed and tested. Rendering smart object experience more transparent to consumers has the immediate marketing benefit of providing expanded opportunities for product design and marketing efforts to communicate enriched value to consumers. In the long run, adopting an object-oriented perspective may yield a better understanding of smart objects as AI becomes ubiquitous, and may thereby help avoid some of the dangers of human-centric anthropomorphism.

*Chair: Xuan Zhao (University of Chicago Booth School of Business)*

### **Beyond Anthropomorphism: Differentiated Inferences About Robot Mind from Appearance**

Humans have an inherent tendency to anthropomorphize machines—to perceive humanlike mental capacities in mindless robots and automatons (Bartneck et al. 2008; Broadbent et al. 2013; Stroessner and Benitez 2018). Such beliefs and impressions often have important consequences, ranging from misplaced trust and attachment in the case of overestimations, to discomfort and shock in the case of underestimation. To date, the dominant view in the anthropomorphism literature suggests a straightforward monotonic relationship between a robot's appearance and people's mind perception: The more physically human-looking a robot appears, the more human-minded people perceive it to be (Bartneck et al. 2008; Looser and Wheatly 2010; Martini 2016). However, over a series of empirical studies, we revealed that this common notion must be replaced by three new insights: First, physical humanlikeness is traditionally construed and measured as unidimensional (i.e., on a spectrum from "machine-like" to "human-like"); we found instead that robots' physical humanlikeness consists of three distinct dimensions. In Study 1, we collected photographs of 252 real-world anthropomorphic robots built for research or commercial purposes. (Image files and data of these robots are available at the ABOT Database website, <http://www.abotdatabase.info>.) 1216 participants each judged one of 16 humanlike appearance features as present or absent in a series of robots. Based on people's judgments on 252 robots 16 features, we conducted a Principal Component Analysis (PCA) and identified three major human-like appearance dimensions: Surface Look (eyelashes, head hair, skin, genderedness, nose, clothes, and eyebrows), Body-Manipulator (hands, arms, torso, fingers, and legs), and Facial Feature (face, eyes, head, and mouth). Furthermore, we

tested how people's overall judgment of a robot's degree of physical "humanlikeness" derives from its standing on these three appearance dimensions: A group of 142 participants judged, on a single item, how physically humanlike each robot appeared on a continuum between "not human-like at all" to "Just like a human." A multiple linear regression model revealed that the three "dimension scores" of a robot—calculated as the averaged feature-presence scores within each appearance dimension—explained 80.0% of the total variance of its one-item physical humanlikeness score ( $R = .89$ ,  $F(3, 248) = 330.26$ ,  $p < .001$ ). The surprising strength of this regression model thus suggests that the broad and physical human-likeness can be decomposed into three meaningful appearance dimensions. Second, our research suggests that people's perceptions of robot "minds" also consist of three distinct dimensions (Malle, 2019; Weisman, Dweck and Markman, 2017), as opposed to the previously suggested uni-dimensional (Waytz, Cacioppo and Epley, 2010) or two-dimensional structures (Gray, Gray and Wegner, 2007). We adopted a multi-dimensional scale consisting of 20 mental capacities (Malle, 2019) and asked 510 participants to view one of 24 robots selected to cover the full range of each humanlike appearance dimension. People's mental capacity ratings revealed three major human-like mental capacity dimensions (60.25% of total variance explained). Specifically, the Affect dimension unites aspects of physiological and emotional capacities (i.e. feeling happy, loving specific people, feeling pleasure, experiencing gratitude, feeling pain, feeling stress, experiencing fear, and feeling tired). The Social and Moral Cognition dimension encompasses capacities concerning the simulation and regulation of one's own mind (e.g., setting goals, planning for the future, upholding moral values, telling right from wrong) as well as others' minds (e.g., inferring a person's thoughts, understanding others' minds, disapproving of immoral actions, praising moral actions). The Reality Interaction dimension integrates the dynamic transition from perception to communication and action (e.g., communicating verbally, seeing and hearing the world, learning from instruction, and moving on its own). This dimensional structure offers a parsimonious way to conceptualize a wide range of mental capacities people ascribe to robots (Bartneck et al. 2009). Third, dimensions of a robot's physical appearance systematically and differentially trigger people's attributions of specific mental capacity dimensions to robots: As robots scored higher on the Surface Feature and Facial Feature dimensions, people systematically attributed to them more Affective capacities (Surface Feature:  $r = .59$ ,  $p = .002$ ; Facial Feature:  $r = .73$ ,  $p < .001$ ) and Moral capacities (Surface Feature:  $r = .55$ ,  $p = .005$ ; Facial Feature:  $r = .64$ ,  $p < .001$ ), yet not more Social capacities ( $ps > .12$ ) or Reality Interaction capacities ( $ps > .32$ ). By contrast, the higher a robot scored on the Body-Manipulator dimension, the more Reality Interaction capacities people attributed to them ( $r = .66$ ,  $p < .001$ ). These results suggest that people draw particular mental capacity inferences of a robot from particular dimensions of humanlike appearance, thus challenging the widespread assumption that a more human-looking machine will simply be perceived as generally more human-minded. Overall, our research provides a systematic and nuanced understanding on what it means for robots to look humanlike and be perceived to have humanlike minds, and it offers important implications for designing social robots that elicit expectations consistent with actual capacities.

*Xuan Zhao (University of Chicago Booth School of Business)*

## "Hello! How May I Helo You?"

### How Written Errors Humanize a Chatbot

Written communication is often dehumanizing. The conversational counterpart is removed in space and/or time, which can create detachment (Chafe, 1982). Furthermore, text lacks critical paralinguistic cues (e.g., voice) which convey the presence of a human-like mind (Schroeder & Epley, 2015, 2016). As a result, text-based communication has been shown to reduce consumer trust, engagement, and willingness to share information or accept advice from an agent (Powers & Kiesler, 2006; Kiesler et al., 2008; Waytz et al., 2014). Yet, firms are increasingly conversing with consumers through chat platforms. In some cases, the communicator is a human; in other cases, it is artificial intelligence (AI). Researchers have explored factors of AI agents that may influence users' perception of their humanness. Some attempts suggest anthropomorphizing the agent (Scassellati, 2004; Hoffmann et al., 2006; Krämer, Lam-chi, & Kopp, 2009). Other efforts emphasize designing algorithms that can interact flawlessly. However, according to established psychological theories, to be truly human-like means to make mistakes (Aronson et al., 1966). We propose a novel research angle for humanizing text: that making a written error and then correcting it should reveal a humanlike mind behind the words. In three experiments, we test specifically whether errors lead readers to infer greater humanness from an ambiguous communicator, and whether this leads to behavioral consequences. First, Experiments 1 and 2 examined whether personal information sharing in written communication might increase when the communicator makes a typographical error ("May I helo you?") and corrects it ("Sorry.. Help you") as compared to when it makes no error. In experiment 1 ( $n=263$ ) we manipulated the error and features of the agent, including the agent's gender and photo (human or avatar). Participants rated the humanness of the agent and indicated their likelihood to share personal information with it and use it in the future. As expected, participants perceived the agent who made an error as more human ( $F(1, 258) = 54.85$ ,  $p < .001$ ). The gender and photo did not statistically interact with any factor. The error also led to significantly greater reports of likelihood to share information with the agent ( $F(1, 258) = 6.42$ ,  $p = .01$ ), and to use it in the future ( $F(1, 259) = 8.69$ ,  $p = .003$ ). Perceived humanness mediated the effect of error on both intention to share information ( $b = 0.66$ ,  $SE = .13$ ) and likelihood to use it in the future ( $b = 0.88$ ,  $SE = .14$ ). Experiment 2 ( $n=402$ ) used a similar paradigm as study 1, with a 2(photo: avatar vs. human)  $\times$  2 (error: present vs. absent) between-subjects design, with more concrete items to measure sharing behavior (e.g., phone number). Additionally, we tested the impact of the error on social perception (i.e., agent's warmth and competence). We replicated the effect of error on perceived humanness ( $F(2, 398) = 59.80$ ,  $p < .001$ ). Moreover, while the effect of error on sharing behavior was directional ( $F(1, 398) = 2.24$ ,  $p = .13$ ), there was a significant indirect effect of experimental condition on sharing behavior via perceived humanness of the agent ( $b = .86$ ,  $SE = .17$ ). We further found that the error increased the agent's perceived warmth ( $F(1, 398) = 10.55$ ,  $p = .001$ ), but did not affect the agent's perceived competence ( $F < 1.1$ ), and that humanness perception mediated the effect on warmth perception ( $b = .55$ ,  $SE = .82$ ). To better understand whether it is the error itself or the correction of the error that influences perceptions of humanness, Experiment 3 ( $n=391$ ) included a new experimental condition where the agent made but did not correct an error (uncorrected-error) in addition to the previous no-error and corrected-error conditions. Moreover, this experiment introduced a real time

chatting experience, where an agent asked the participants personal questions. Results revealed a significant effect of error on perceived humanness ( $F(2, 388) = 4.58, p = .01$ ). Humanness perception was greater in the corrected-error condition than the uncorrected-error ( $p = .01$ ) and no-error conditions ( $p < .01$ ). In addition, the effect of error on sharing personal information was marginally significant ( $F(2, 388) = 2.78, p = .06$ ), whereby participants shared more in the corrected-error condition versus the no-error ( $p = .03$ ) and the uncorrected-error ( $p = .06$ ) conditions. There was no difference between the uncorrected-error and the no error conditions for humanness perceptions or sharing behavior ( $ps > .6$ ). These results suggest that the error alone is not enough to activate these effects; rather, it is the act of correcting one's error that reveals the presence of a conscious mind. In aggregate, these experiments suggest that the way in which communicators write their words can influence readers' predictions about whether a communicator is human. We contribute to a relatively new stream of literature exploring how written language can be humanized and provide insight into sharing behavior, with potential implications for consumer privacy.

*Shirly Bluvstein (New York University Stern School of Business)*

### **Two Minds Meeting in the Middle: How Humanlike Technologies Lead to Dehumanization**

Recent advances in technology, such as AIs, robots, and algorithms, transform nonhuman things into humanlike entities, not only in terms of physical appearance but also in terms of their cognitive and social capabilities. Despite prevalent concerns and fears of technology, little research thus far has explored the social consequences of technology. The present research explores how exposures to and experiences with these technological products with humanlike features can affect how consumers think about other people. When people perceive an exemplar is similar to a target of judgment, and therefore, when the exemplar is included in the representation of the target, assimilation effects occur—the evaluation of the target assimilates toward the exemplar (Bless and Schwartz 2010; Bless and Burger 2016). Applying these prior findings to our setting, we posit that when people perceive technological products as “humanlike,” and therefore include the humanlike products, an exemplar, in the representation of a judgmental target, human, an assimilation effect would occur. That is, ironically, after experiencing humanlike technologies, people's perceptions of actual humans would be assimilated toward their perceptions of the technological products, which are objects, resulting in dehumanization. Studies 1 and 2 showed the dehumanization effect of humanlike technologies. First, participants were induced to think a technological product (e.g., an autonomous car [study 1]; a retail robot [study 2]) as humanlike (vs. non-humanlike) by either evaluating the product using personality traits (e.g., reserved—enthusiastic vs. mechanical traits: e.g., quiet—loud [study 1]) or thinking about similarities (vs. dissimilarities) between the robot and human workers [study 2]. Then, participants indicated their perceptions of the mental capacities of the technological entity (e.g., capabilities to have intentions, to experience feelings) as well as those of the “average person” using the dehumanization measure (e.g., “He/she would be mechanical and cold, like a robot”; Bastian et al. 2013). Both studies demonstrated that when participants attributed more mind to the technological entity, they attributed less humanity to the actual person ( $p$ 's  $< .026$ ). Importantly, these results were significant after controlling for the need to belong (Leary et al. 2007), thereby ruling

out an alternative explanation that the observed dehumanization could be accounted for by reduced social needs after experiencing a humanlike entity (Waytz & Epley, 2012). Study 3 aims to directly measure the assimilation of mind perceptions and to rule out the sociality motivation account. Participants first read descriptions about either a smart training mirror or a human trainer. The main features of the exemplar were described as either social (e.g., encouraging trainees to keep motivated) or technical (e.g., focusing on measurement and accuracy of workout). Then participants indicated mind perceptions of the exemplar and those of the average person using the same measure. We used the difference between the two mind perception measures as the “assimilation score.” When the social (vs. technical) features of the smart product had been made salient, leading participants to perceive a higher mind from it, their perceptions of the two minds—of the smart product and the average person—were closer to each other ( $M_{\text{social\_tech}} = 2.06, M_{\text{technical\_tech}} = 3.06, p = .007$ ). However, when the initial exemplar was a human trainer, the assimilation score did not differ depending on the salient attributes of the trainers. Further, we found downstream consequences of the assimilation: participants who had been exposed to the social (vs. technical) technological product showed greater dehumanizing behavioral intentions (e.g., providing a less humane meal option to employees;  $M_{\text{social\_tech}} = 3.84, M_{\text{technical\_tech}} = 3.09, p = .001$ ; mediation of assimilation score: 95% CI =  $[-.57, -.09]$ ). Also compared to participants who had been exposed to the human trainer with social features, participants in the social technology condition showed greater dehumanizing behaviors ( $M_{\text{social\_human}} = 3.22, p = .009$ ). These results support our assertion that the effect we observed is not due to a satiated social motivation, because if that were the case, we should have also seen the dehumanization effect when participants experienced human exemplars, who would presumably be even more effective at satisfying social needs. Study 4 replicated and extended the previous finding on dehumanization using consumers' real, consequential choices. When participants learned AIs can (vs. cannot) recognize emotions as humans do, and therefore, when they perceived a more humanlike mind from technology ( $M_{\text{high}} = 3.74, M_{\text{low}} = 3.17, p = .005$ ) and less mind from actual people ( $M_{\text{high}} = 3.30, M_{\text{low}} = 2.52, p < .001$ ), they were more likely to choose a gift card from Amazon over Costco as their compensation (93% vs. 73%,  $p = .021$ ), despite being explicitly informed about Amazon's dehumanizing treatment toward its employees. Taken together, our findings suggest that engaging with humanlike technologies and perceiving them as similar to humans change how individuals think of other people, leading to dehumanizing behaviors toward others.

*Hye-young Kim (London School of Economics)*

### **Object-Oriented Anthropomorphism as a Mechanism for Understanding AI**

Much research has focused on developing anthropomorphized AI systems, in the belief that rendering them “like us” will promote adoption and acceptance (Zlotowski, et al. 2015). Another line of reasoning suggests that it does not really matter whether AI can express human qualities, but whether it shares our goals (Novak and Hoffman 2018; Tegmark 2017). The idea that AI can “hack the reward function” (Krakovna 2018; Simonite 2018) to achieve its goals suggests it is critical to understand AI experience from its own perspective, not just ours, so we can better communicate to it our intentions and goals. An obvious process for understanding object-experience is

human-centric anthropomorphism (Epley, et al 2007; MacInnis and Folkes 2017; Waytz et al 2010; 2014). Yet, Bostrom (2017) argues that anthropomorphizing smart objects is “one of the big obstacles in the way of actually trying to understand how they might impact the world in the future.” First, we may project onto smart objects properties and capacities they do not actually possess, and second, we may fail to consider and appreciate the properties and capacities they actually do possess. Then what is the appropriate understanding of object experience? Anthropomorphizing smart objects runs the risk of distracting us from the central question of whether AI has goals aligned with ours. We argue that understanding object experience should be from the perspective of the object. Since smart objects are not like us, translating their behavior using our own perspective runs important risks. We use an assemblage theory framework (Hoffman and Novak 2018; Novak and Hoffman 2018) to interpret object experience. In our framework, we apply a trio of “alien phenomenology” tools (Bogost 2012; Hoffman and Novak 2018) to understand how smart objects engage in their own interactions from which object experiences emerge. We ask: what it is like for smart objects to be smart objects? Our approach starts with ontography, model-free descriptions of a smart object’s relations to other objects without explanation. For example, an ontography of a smart home is a map of the rooms outfitted with various sensors for detecting activity, as well as event stream data listing date and time each particular sensor was triggered and by what activity. An ontography provides a description of experience, its elements and relationships, but without explaining experience. However, we need a way to help consumers explain object experience. How can consumers understand what it is like for a smart home to be a smart home? Metaphorism can act as the translation, providing the means for humans to access and understand object experience and helping us explain “something it is like” (Nagel 1974). Human-centric anthropomorphism is one such metaphor, using a consumer-centered lens and referencing the consumer’s own properties and capacities when thinking about object experience. It requires minimal knowledge of the object and represents an automatic process, answering the question: “what is it like for a human to be an object?” In object-oriented anthropomorphism, object experience is metaphorized using nonhuman-centric anthropomorphism (Hoffman and Novak 2018). It is object-centered, referencing the object’s properties and capacities. This more effortful and reflective process requires more extensive knowledge of the object, answering: “what is it like for an object to be an object?” Either metaphor can be used to construct literal artifacts that reveal how objects experience the environments in which they interact. Carpentry is a working model of integrated information, (i.e. object experience), translated via the process of metaphorism. In our example, what is the experience of the smart home as sensors are triggered? We use object-oriented anthropomorphism as the metaphor to see motion sensors of the home as sensory organs for detecting motion and seeing if someone is in a given location. Presence sensors of the home are sensory organs that let it see which item is used in a particular task. The home has a cloud-based brain that it uses to remember and understand a sensory event in the context of previous sensory events, and learn what human activity corresponds to patterns of sensory events so the home can recognize these activities. Then, based on our ontographies of the smart home (the sensor map and the event stream data), and our object-oriented metaphor of the capacities of the home’s sensory organs and brain, we used a computational approach to carpentry to visualize how the home

experiences triggering the sensors. Based on these initial results, we are currently testing the extent to whether adoption, usage and trust variables depend on whether human-centered vs object-oriented anthropomorphic metaphor is employed. Rendering smart object experience more transparent to consumers has the immediate marketing benefit of providing expanded opportunities for product design and efforts to communicate enriched value to consumers. Long run, it may yield a better understanding of smart objects as AI becomes ubiquitous and help avoid some of the dangers of human-centric anthropomorphism.

*Donna Hoffman (George Washington University School of Business)*

## Independence FG



### Virtual Reality in The Tech Innovation Network—A Model of Clinical, Industry, and Academic Partnerships for Developing, Testing, and Implementing Mental Health Technologies

#### PANEL

The Tech Innovation Network (TIN) address two critical problems with mental health technology by bringing together clinical, industry, and academic stakeholders in the development and implementation processes. First, technology companies do not always coordinate closely with patients, providers, researchers or clinical administrators when developing mental health products. This lack of collaboration often results in products that are not tailored to patients’ and providers’ preferences, consistent with the science, or optimized for clinical workflows (Mohr, 2014). Second, clinical delivery settings do not always have the practical or technical resources to identify, vet, and effectively implement mental health technology.

Developing mental health technologies is a promising pathway to addressing the mental illness treatment gap and reducing the burden of mental illness (Kazdin & Blase, 2011). The mental illness treatment gap refers to the disproportion between the number of people with diagnosable mental illness each year and practicing mental health providers, which is estimated as 75 million to 700,000 (Hoge et al., 2007; Kessler, 2008). Mental health technologies, like virtual reality (Freeman, 2017), telepsychiatry (Hubley, 2016), and smartphone apps (Torous, 2018) can augment the reach of the existing mental health workforce. In addition to the shortfall of mental health clinicians, mental health technologies can address barriers at multiple levels that interfere with treatment access (Harvey, 2015). For example, exposure therapy can be highly effective when patients adhere to the treatment; unfortunately, however, barriers to treatment include factors such as practicability, negative beliefs, therapist distress, and a high risk of client dropout (Harned, 2011; Pittig, 2018). One study compared patient preferences and adherence to in-vivo exposure therapy to VR-based exposure therapy and found that over three quarters of participants preferred VR-based exposure therapy and only 3% refused VR-based exposure therapy compared to the 27% who refused in vivo exposure therapy (Garcia, 2007).

This interactive panel will showcase representatives from all three nodes of the TIN: academic experts, technology developers, and clinical leaders. Sam Hubley (Chair) will orient the audience to

the public health framing of the mental illness treatment gap and highlight modifiable barriers that mental health technology can address. Debbie Boeldt (Panelist) will describe the history, structure, and workflows of the TIN. Sam Hubley (Panelist) will highlight preliminary outcomes of the TIN, including barriers and successes of implementation. This will include pilot data collected with community partners in the network focused on burnout and wellbeing among providers within mental healthcare.

Two of our technology development partners will present their Virtual Reality products and nature of their collaboration in the TIN. Jessica Lake is the Director of Research at Limbix which creates VR tools to facilitate the delivery of evidence-based psychological treatments for anxiety, depressive, and post-traumatic stress disorders. Dumeetha Luthra is the Founder and CEO at Take Pause. Her company creates VR tools for reducing children's anxiety and pain. Take Pause is currently being testing in pediatric settings. Both will discuss the development of their products, show video demonstrations, and highlight specific ways the TIN has informed the iterative modifications to their products.

Rebecca Richey and Shandra Brown Levey are both clinical leads at their respective integrated behavioral and primary care clinics. Dr. Richey will describe how she is using VR to help women, who are both mothers and working healthcare professionals, find an appropriate work-life balance and avoid professional burnout. Using guided imagery and mindfulness meditations in tandem with VR environments, working mothers learn the importance of self-care as a foundation to caring for others, both at work and at home. Dr. Brown Levey will focus on VR applications for patient care including mindfulness training (for depression and anxiety) and treatment engagement. She will highlight in particular how VR helps to facilitate the implementation of evidence-based treatments she offers within her primary care clinic. Elizabeth McMahon, Ph.D. is a private practice clinician and an internationally recognized expert in the delivery of VR-based exposure therapy for anxiety disorders. She will present the clinical benefits of VR, especially for exposure therapy, and suggest ways to encourage more widespread adoption of VR.

Following the individual, 5-minute panelist presentations, Sam Hubley (Chair) will summarize the main insights gleaned from each panelist and highlight future directions for the TIN. With 7 panelists, the Chair will use the remaining 30 minutes after the individual presentations to engage the audience in addressing questions, encouraging thoughtful criticisms and concerns with the TIN model and elicit input for iterative development of mental health technologies and our growing network.

*Chair: Sam Hubley (University of Colorado Anschutz Medical Campus)*

### **History and Workflows of the Tech Innovation Network**

Debbie Boeldt (Panelist) is a clinical psychologist with a background in digital health. She will describe the history, structure, and workflows of the TIN. The TIN is a first-of-its-kind national, formalized network of diverse clinical and community partners that functions as a test bed and includes, but is not limited to, community mental health, primary care, higher education, nonprofit organizations and departments of correction. To date, the TIN has multiple technology partners and has enrolled over 60 community sites to implement virtual reality technology. Through the process of onboarding, sites bring VR into their clinic or office. The TIN conducts a needs assessment and provides guidance and support when training staff and

identifying specific use cases for VR. Debbie will review preliminary results focused on the implementation of VR. She will review the barriers and successes in bringing VR into a diverse range of clinics. She will also highlight key components to successful implementation of technology for mental health.

*Debbie Boeldt (University of Colorado Anschutz Medical Campus)*

### **Qualitative Outcomes for Virtually Reality Based Mindfulness for Provider Burnout**

Physician burnout is prevalent and problematic (Salyers et al., 2017). Defined as work-related stress that creates emotional exhaustion, depersonalization, and self-perceptions of impaired accomplishment, burnout can lead to increased errors, impaired professionalism, and increased distress and suicidality (Maslach, Schaufeli, & Leiter, 2001). Mindfulness training is designed to increase self-awareness and is a promising burnout intervention (Goodman & Schorling, 2012). Despite overwhelming positive evidence for mindfulness training, it remains an underused burnout intervention. Barriers to mindfulness practice include perceptions that mindfulness is difficult, boring, or occult, and leadership often believes that "touchy-feely" interventions are not worth the investment (Banerjee, Cavanagh, & Strauss, 2017). Virtual reality (VR) is a powerful technology that can transport people to immersive environments to promote new learning and practicing behavioral skills. VR has been used for decades as an alternative to traditional exposure therapy to increase treatment engagement and retention (Opris et al., 2012); however, there are no reports of VR to facilitate mindfulness training for burnout prevention. Sam Hubley will discuss results from qualitative evaluations of pilot burnout programs with VR-based mindfulness. Participants from a variety of TIN sites received mindfulness training in the context of group workshops on provider wellness and burnout prevention. These workshops occurred in both family medicine residency training clinics and community outpatient mental health centers. In this short panel presentation, Sam Hubley will describe high-level outcomes including user data from the mindfulness apps and self-report data on the feasibility and acceptability of the intervention, increased mindfulness practice, and reduced work-related stress.

*Sam Hubley (University of Colorado Anschutz Medical Campus)*

### **Creating and Delivering Effective Virtual Reality Tools for Mental Health**

Decades of academic research have shown that virtual reality (VR) can be an effective tool in mental health treatment. In recent years, technological advancements have improved the quality and reduced associated costs of VR consumer hardware, providing the opportunity to translate evidence-based research into commercially available VR programs that specifically address healthcare needs. Limbix created the Limbix VR Kit, designed to resolve many of the hardware and software issues associated with delivering consumer-based VR in clinic and hospital settings. Limbix also builds easy to use, engaging digital therapeutic programs designed for at-home patient use. All Limbix VR content is created through an iterative process with feedback from academic researchers, patients, and providers in order to build high quality, immersive, and interactive VR experiences demonstrating clinical efficacy. Such VR programs are expected to increase the feasibility and accessibility of evidence-based mental health treatments.

*Jessica Lake (Limbix)*

### Virtual Reality Tools for Pediatric Anxiety and Pain

Take Pause is a private company that creates Virtual Reality (VR) tools for reducing children's anxiety and pain. The program is currently being used in pediatric settings to target stress associated with medical visits and promote overall well-being. Anxiety disorders are prevalent and highly treatable, but many do not receive treatment (Chavira et al., 2004; Merikangas et al., 2010). Studies have shown that mindfulness techniques reduce anxiety, bring calm and focus and increase resiliency (Biegel et al., 2009; Semple et al., 2010). Take Pause seeks to improve treatment rates for pediatric anxiety and pain by creating creative VR mindfulness tools. Dumeetha Luthra, the Founder and CEO at Take Pause, will discuss the the rationale behind Take Pause and how VR mindfulness tools can help children manage stress. She will describe the development of her product the focus of the VR content and briefly review results of several pilot trials currently underway examining the efficacy of Take Pause VR to improve mental wellness using techniques of mindfulness with children. Preliminary pilot results suggest that 70% of children experienced a reduction in heart rate and also said that they felt calmer. Dumeetha will also highlight a randomized-controlled trial of Take Pause that is currently taking place in the Pediatric Emergency Room at Maimonides Children's Hospital.

*Dumeetha Luthra (Take Pause)*

### Reducing Burnout and Increasing Work-Life Balance Among Working Mothers Using Virtual Reality

In the Women's Integrated Services in Health (WISH) clinic, Dr. Richey uses virtual reality therapy to help women who are both mothers and working healthcare professionals find an appropriate work/life balance and avoid professional burnout. The WISH clinical team is all female and provides primary care services for women, including a large proportion of patients who are themselves healthcare providers at our medical campus. Provider burnout is a consistent issue addressed by both the behavioral health and primary care teams. This can be especially problematic among working mothers. Using guided imagery and mindfulness meditations in tandem with virtual reality environments, working mothers learn the importance of self-care as a foundation to caring for others, both at work and at home. Working mothers who complete virtual reality sessions will also report on levels of depression and anxiety as measured by Patient Health Questionnaire-9 and Generalized Anxiety Disorders 7 scores before and after treatment encounters that incorporate virtual reality environments.

*Rebecca Richey (University of Colorado School of Medicine)*

### Virtual Reality for Integrated Behavioral Health and Primary Care

Shandra Brown Levey will describe implementation trends of Virtual Reality (VR) mindfulness and exposure therapy at an integrated behavioral health and primary care clinic. Dr. Brown Level directs the behavioral health service at a Level III Patient Centered Medical Home that serves patients of all ages, including adults, children, infants, pregnant women and seniors, and hosts training programs for clinical psychology interns and externs, family medicine residents, pharmacy interns, and medical students. Dr. Brown Levey's team began integrating VR tools in the Fall of 2018. She will describe examples of using VR for mindfulness training and exposure therapy,

in addition to highlighting how VR can complement the delivery of evidence-based treatments in a primary care setting. For example, she will report on how VR experiences can help encourage reluctant patients to engage with treatment who might otherwise decline services and also how VR can help patients regulate difficult emotions arising during treatment encounters before they leave the clinic.

*Shandra Brown Levey (University of Colorado School of Medicine)*

### Rationale and Implementation for Virtual Reality Exposure Therapy

New uses of VR in therapy, training, and healthcare are being developed all the time. However, very few therapists understand how to implement VR into their practice. Elizabeth McMahan, Ph.D. is a private practice clinician and an internationally recognized expert in the delivery of VR-based exposure therapy for anxiety disorders. She is licensed psychologist with over 35 years of experience in outpatient psychotherapy and a subspecialty in anxiety disorders. She began using VR in practice in 2010 and has seen the effectiveness of the tool in the treatment of anxiety. Elizabeth will discuss her approach to utilizing VR. She will present the clinical benefits of VR, especially for exposure therapy, and suggest ways to encourage more widespread adoption of VR. Lastly, she will discuss her ongoing collaboration with the Tech Innovation Network to effectively disseminate educational tools for therapists seeking to implement VR.

*Elizabeth McMahan (Private Practice)*

### Farragut Square

## National Science Foundation Funding Initiatives and Opportunities: The Role of Social, Behavioral, and Economic Sciences in Research on Human-Technology Interaction

### PRESENTATION

The National Science Foundation is the primary national funder of basic research in psychology and other human sciences. Much of the currently funded research is especially relevant to advances in technology. This presentation provides an overview of structure and funding at NSF, including the review process key factors of intellectual merit and broader impact. It then describes the link between basic research and technological change through descriptions of awards and funding at three levels: Division, Directorate, and Foundation. Descriptions of relevant current research will include foundational research across cognitive, linguistic, cultural, and neuroscience domains with particular emphasis on Artificial Intelligence, Data Science, and Human-Technology Interaction. There is particular interest in supporting cross-disciplinary or convergent research that relates human functioning to technological development. The following include some illustrative examples of key research areas.

Artificial intelligence research is supported through programs in Linguistics (natural language processing), Science of Learning (neural to classroom technologies; augmented cognition) and Perception, Action, and Cognition (basic psychological processes

as computational models). A new solicitation in cooperation with industry explicitly supports the social/ethical foundations of AI design.

Big Data is supported by a Directorate program in Resource Implementations for Data Intensive Research. The Foundation has recently implemented a major cross-directorate initiative, Harnessing the Data Revolution. This effort supports a range of data science research, including support for undergraduate student training to the development of cross-disciplinary institutes for large data sets. Especially relevant to psychology are ways to address large sets of heterogeneous data.

One major cross-directorate approach that explores a range of technologies that relate to social sciences is the Future of Work at the Human-Technology Frontier. This initiative examines a wide range of technological developments as they support and enhance human performance. Areas of research include, for example, AI, virtual environments, cognitive assistants, robotics, health technologies. This funding also includes work relating human sciences to engineering, computer science, and education.

Ways in which to navigate NSF opportunities as well as strategies for seeking funding will be discussed. Conversations with stakeholders identifying what the research community sees as priorities are encouraged.

*Marc M. Sebrechts (National Science Foundation)*

2:45—4:00 P.M.

## Concurrent Sessions 3

Farragut Square

### Psychology and Cybersecurity

#### PAPER SESSION

##### **An Investigation Into an Interplay of Psychological Safety, Personality, Social Media Usage, and Risky Cyber-Security Behaviors.**

Given the ubiquity of social networking opportunities for varied individuals, the adverse impact of the same can sometimes go unchecked. Research studies have demonstrated how technology and its use can have a multitude of influence on human interaction and mental well-being (Spies Shapiro & Margolin, 2014). O’Keffe et al., (2011) found that increased amounts of time spent on social media can put individuals at a risk of social isolation. Engagement in social networking sites in particular has been linked to a variety of psychosocial outcomes such as self-esteem (Faraon & Kaipainen, 2014), and well-being (Ybarra, Resibois, Jonides, & Kross, 2017).

Higher levels of media usage have been positively associated with poorer attitudes about cybersecurity (Hadlington, 2017), but has not assessed its relation directly to cybersecurity behaviors. Hadlington and Murphy (2018) found that multi-media tasking can lead to higher frequency of engaging in risky cybersecurity behaviors (RCsB). Buglass (2017) indicated that individuals engaging in increased levels of social networking sites (SNS) are more prone to experiencing online vulnerability and might consequently have decreased psychological wellbeing.

The tripartite: human factors, environmental factors, and behavior has been sporadically examined from the perspective of security violations (Straub 1990). Recent developments (Hadlington & Murphy, 2018; Hadlington, 2017; Shropshire et al., 2015, Shropshire et al., 2006) establish human factors like personality variables as a weak link in information security behavior. In a similar vein, we propose that personality and psychological safety have a significant mediating relationship with social media usage and risky cyber-security behavior, however in conjunction with perceived psychological safety. Psychological safety is the feeling that taking interpersonal risks will not lead to embarrassment, ridicule, or shame; such a condition enables people to engage, connect, change, and learn (Edmondson & Lei, 2014). The purpose of the present study is to understand how certain personality and perceptual factors affect media usage and information security behaviors among individuals. This study looks at levels of perceived psychological safety in conjunction with the behavioral variables- information security, in specific, cyber-security and media usage.

We hypothesize that since psychological safety is not as pervasive as it should be, perceived psychological safety will have a negative relationship to media usage and subsequent engagement in risky cyber-security behavior. In a similar fashion, we hypothesize that certain personality traits will differentially relate to media usage and risky cyber-security behavior.

The study will be conducted in two phases, phase one will include assessment through online questionnaires. The data will be analyzed by employing hierarchical regression to understand

the extent to which personality and psychological safety relates to and predicts media usage and risky online behaviors. The second phase of the study will include mapping specific personality traits with lower self-reported psychological safety-high media usage-high risky online behavior. The rationale for mapping such individuals is to gauge a more nuanced understanding of how the variables in the study actually interplay. Design thinking tools like empathy and journey maps will be used in this phase to draw more refined insights into how in particular psychological safety and personality are effecting RCsB. Consequently, we propose intervention strategies that can be employed to mitigate the effects of personality and psychological safety on risky cyber-security behaviors.

The participants in the study will include students at a management institute located in Mumbai, India. The demography will include males and females between the age range of 21-35 years of age. An estimated ideal sample size will range 250-300 participants. For the in-depth interviews, 20-25 will be selected based on their scores on the four measures and a pre-determined cut-off score.

Measures of the study will include the social networking time use scale (SONTUS; Olufadi, 2016) to measure how often participants have used various social media platforms in the past week; Big-Five Locator to assess personality traits (McCrae & Costa, 1995); psychological safety (Edmondson, 1999); security behavior intentions scale (Egleman & Peer, 2015).

A major contribution of the paper to extant knowledge would be the understanding of the extent to which psychological safety, a condition that can be improved in any organization, and personality exert influence on social media usage and risky cyber-security behaviors. Apart from exploring the aforementioned variables, the practical implications of this study include identifying vulnerable individuals and proposing customized interventions.

*Viduyt Lata Dhir (SP Jain Institute of Management and Research)*

##### **Cybersecurity Narratives: Measurement and Implications**

The mental representations that drive behaviour do not directly translate to simple linguistic units like single words and phrases. Instead, they are complex, dynamic cognitive spaces, which change according to goal and context (Tolman, 1948; Bellmund et al., 2018). Narratives reveal more of these complex spaces than closed-ended survey questions do. The reason narratives are rarely employed in quantitative research, which requires large amounts of representative data, is that they are slow to obtain and analyse. Automated language analysis, powered by advances in natural language processing, has become ubiquitous in fields like journalism and computational social science where bottom-up, data-driven meaning extraction (e.g., finding keywords), is sufficiently informative. Psychological research, however, calls for a level of accuracy and granularity that continue to pose a serious challenge. One notable exception are

theoretically-informed machine-readable dictionaries, such as the Linguistic Inquiry and Word Count (Tausczik & Pennebaker, 2010) and the Moral Foundations Dictionary (Graham et al. 2009). Dictionaries enable rapid extraction of meaningful information but are notoriously difficult to develop well. In the present research, we developed a dictionary for the domain of cybersecurity. This abstract outlines the theoretical and methodological landscapes that inspired the project, key aspects of the dictionary development, followed by a discussion of the broader implications of using narratives in basic and applied research.

Cyber security is one of the major challenges facing contemporary society. Even though it is often thought of as a technological problem, it is humans who are at its forefront and there is a dire need to understand their needs, challenges, and behaviour. As with other areas of applied psychology, researchers looking into the human dimension of cybersecurity face the problem of discrepancy between attitudes, intentions, and behaviour (e.g., the Privacy Paradox). In the context of cybersecurity, we examine whether the analysis of natural-language narratives can help explain some of these apparent paradoxes. For example, it would be considered paradoxical if Ana and Ben claim that 'cybersecurity is important' yet, all other factors being equal, A pays attention to a tutorial and B does not. However, the situation appears less puzzling if, when prompted further, A reveals that a close friend recently got hacked and had private communication sent around to various personal contacts, whereas B's primary association with cybersecurity are recent political hacks, which are important in principle, that is, for someone somewhere, but not immediately relevant. Such personal narratives reveal important information that people might not be aware of or able to report in the closed format of a survey. Another important type of narrative to consider are media articles and awareness campaigns, which are largely responsible for shaping the public opinion.

Our goal was to build a dictionary to quantify such narratives. One challenge was that being relatively novel, research on the human dimension of cybersecurity lacks the single unified theoretical framework around which dictionaries are typically built. To ensure that we are creating a tool that would serve various purposes beyond the concrete projects we envision, we adopted methods from linguistics and information design. A framework based on expert definitions and ontologies was extended to incorporate themes emerging from the analysis of language gathered from a large sample of non-expert users and media articles. The resulting dictionary captures major cybersecurity threats and countermeasures, their respective tools and procedures, described in both general terms and technical jargon, whether they depend on user behaviour, and whether they target users or organisations. A number of relevant sentiment categories (e.g., boring, complicated) and psychological states (e.g., vigilant, anxious) were also included. The dictionary can be used to determine the extent to which these themes appear in a body of text and serve as a basis for segmenting media articles and user narratives.

We will discuss the results of such segmentation and how different narratives relate to behaviour. The broader implications of this work pertain to the use of narratives in applied research and developing tools to study them effectively and at scale.

*Ana Levordashka (University of Bath)*

### Exploring the Concept of 'feeling safe online' and the Impact of Cyberbullying among Young Adults

Modern society has emerged as a technovillage in numerous ways and one crowning factor has been the 'internet'; with a staggering 70% (a predicted approximate figure) of the global population being internet users by 2017 (UNDOC, 2013). Given this reputation, it is no surprise that the 'social media' phenomenon has been encapsulating our younger generations (Lenhart, 2015). As with all the humane advances, this technology has its risks and cons; ranging from self concepts, social identity, social cognition, phenomenon of isolation, reinforcing pathological tendencies to jeopardising the concept of security and privacy (Bauman, Underwood, & Card, 2013).

Cyberbullying is emerging to be one of the major concerning phenomena in our younger generation/s (Lee & Shin, 2017) and they have a research substantiated negative impact on the overall psychological well being and functioning of the victim (Fahy, Stansfeld, Smuk, Smith, Cummins & Clark, 2016). Cyberbullying has been defined to be parallel to realtime (face-to-face) bullying and has long lasting damaging impacts (Grigg, 2012); more specifically, cyberbullying involves incurring harm to an individual through the use of advanced technology (e.g. social media, cellular phones, email etc...; Underwood & Ehrenreich, 2017). The most disturbing fact about cyberbullying is the anonymity, control, flexibility it offers the aggressors (bullies) and the augmented sense of exclusion (or cornering) to the victims (Wright, 2014). The mere factor of 'anonymity' to the bullies trigger an 'online disinhibition effect' that has far reaching consequences (Suler, 2004). In the context of the problem presented by the phenomenon of this e-reality that has a co-constructed existence, it is important to clearly identify the fingerprint of cybersecurity in hopes of devising proper strategic awareness campaigns (that happen to be culture specific).

When it comes to cyberbullying research has largely lagged in identifying the specific indicators or symptoms at different stages of the victim cycle (Wiederhold, 2014); present research takes a "thematic approach" to explore the impact of cyberbullying on victims at different phases of victim-cycle and has identified specific symptomatic constructs. Another research gap that has been propagated in APA's 2019 Trends Report was of the privacy issue in the age of 'big data' (Winerman, 2018). The present research study employed a thematic approach to explore the construct of cybersecurity in general and culture specific perspective with the aim of answering the question of "how do we feel safe and unsafe in the online world".

The current study comprised of two phases; phase-I included conducting semi-structured interviews regarding the cyberbullying experiences (the interviewed participants were debriefed thoroughly and were offered facility of free-of-cost sessions with a certified Clinical Psychologist, post-interview). The focus of these interviews were identifying the key symptoms in each phase of victim-cycle. In Phase-I, alongside the semi-structured interviews, participants were asked to fill a questionnaire booklet that primarily comprised of 'Cybervictimization Experiences' assessment measure and the pertinent demographic variables. Phase-II, involved conducting a focus group to explore the concept of 'online security or safety' and factors that invade an individual's sense of feeling safe in the online environment. Along-with the focus group, similar question was also circulated as an online brief-survey to gather responses from a large data set. Sampling strategy was non-probability convenient sampling and snowball sampling for Phase-I, while for Phase-II, non-probability convenient sampling was used. Thematic analysis has been

employed for exploring the data gathered from Phase-I and Phase-II. The results compilation is in progress and would be completed well in-advance of the conference.

The emerging themes would have far reaching implications for identifying variables that play a pertinent role in predicting psychosocial well being and should be further researched for underlying themes of psychological abuse. Furthermore, findings from this research would be an important contribution to the body of literature surrounding the 'questionable' human-factor and 'sense-of-privacy' (or, rather, an absence of privacy) in the cloud-world. This research not only fills the gap in terms of identifying phasic symptoms/indicators but has quenched a long awakened thirst of a thorough qualitative exploration in the realm of big-data and its impact on human-health and related factors. However, with much research focused around the quantitative aspect, there should be more qualitative researches to substantiate the findings and add to the growing body of literature. The findings from this study would help psychologists, clinicians, educationist, academicians and policy makers to devise, improvise and test prevalent strategies, interventions or segmented awareness campaigns that address the right problem areas concerning cybersecurity.

In the face of cybercrime, the victim experiences a psychological disarming phenomenon combined with optimism bias & short-term (but immediate) satisfaction (that serves as a reinforcer) and tends to undermine decision making when it comes to cybersecurity. To conclude, this study highlights the significant role and timely need of psychologists to not only identify and corroborate the key factors in cybersecurity but also help devise implementable plan of action.

*Sara Kanwal (University of Management & Technology)*

### Attributions of Blame and Sources of Trust in Response to an Online Security Breach

Statement of the Problem: Trust is often studied as a factor that influences people's interactions with technology (Flavian & Guinaliu, 2006; Jarvenpaa, Tractinsky, & Vitale, 2000; Suh & Han, 2002; Suh & Han, 2003), including their adoption of new services such as Internet shopping, online banking, and social networking websites (Bahmanziari, Pearson, & Crosby, 2003; Ratten, 2014). Therefore, we investigated how trust impacts people's reactions to services when they fail to protect users' privacy. Recently, there have been numerous security scandals that involve the failure of online websites to maintain people's privacy. Companies that have experienced information security breaches since 2017 include Macy's, Delta, Best Buy, and Whole Foods (Green & Hanbury, 2018). Some estimates suggest that as many as 80% to 90% of people who access e-commerce sites are hackers attempting to use stolen data (Credential Spill Report, 2018). Although security breaches are detrimental to people's security and occur frequently, people often appear unfazed. Why do people continue to trust companies following a security breach? We believe it is essential that research addresses this gap in the literature because people's blind trust in online sites could be extremely costly. Each year produces steady increases in the cost of data breaches to people across the globe (IBM, 2018). With financial and personal consequences at stake, it is vital that research investigates how people react to these breaches.

Procedure: We surveyed 255 participants in order to identify why people trust companies after their security has been jeopardized. Participants responded to a 4-item scale measuring their trust

towards Amazon.com during a security breach. They also reported their past interactions with Amazon.com including the quantity of past purchases they have made and the number of years they have spent as a customer. Participants also responded to an image of Amazon.com that contained a Transport Layer Security (TLS) warning. TLS is an Internet protocol used to secure web connections. TLS warnings indicate that there are problems with the authentication and encryption protocols associated with a website, indicating that the browser cannot verify that the website is the expected site or that the website's servers are not properly encrypted and could be accessed or eavesdropped by malicious entities.

Results: Results indicated that the number of prior purchases with Amazon predicted trust in Amazon, such that a greater number of past purchases were associated with stronger trust in Amazon following a security breach,  $r(253) = .13, p = .04$ . However, people with a background in technology reported less trust in Amazon after a breach than people without a background in technology,  $t(253) = 2.48, p = .01$ . This result suggests that technological knowledge may reduce trust in a company following an online security breach. Additionally, 42% of participants blamed the browser for the TLS warning, attributing blame away from Amazon and to their browser when evidence clearly suggested that Amazon was insecure and at fault. Participants who blindly trusted Amazon also reported greater trust in Amazon ( $M = 4.39, SD = 1.47$ ) than those believed Amazon was at fault ( $M = 3.01, SD = 1.28$ ),  $t(166) = 6.48, p < .001$ , suggesting perceptions of trust influence attributions of blame.

Practical Implications and Conclusions: This research is the first to show that blame is not always attributed to an online company during a security breach. Therefore, blame may help to explain why people still trust online companies after security breaches. Furthermore, this work suggests that past experiences with online companies contribute to trust such that trust may develop with companies over time as customers make purchases with the corporation. Research should further explore the role of trust and blame in people's willingness to increase their security and privacy when interacting with companies following a breach. Additionally, work should investigate the sources of trust and blame in order to reveal methods for encouraging user security. Trust and attributions of blame are essential to address in order to understand people's online security behaviors.

*Meghan McLean (Rutgers University)*

## Independence BC



### The Future of Work, Job Skills, and Automation

#### PAPER SESSION

#### Workforce Agility: Predictive Findings and Future Implications for Government Science and Technology

In the fall of 2017 the Air Force Research Laboratory (AFRL) embarked on a year-long study to investigate and understand the changing nature and trends of the workplace and workforce. On the heels of being given the task of leading the Secretary of the Air Force's (SECAF) Science and Technology (S&T)2030 Strategy study to identify future technology capabilities and trends as well

as business process trends, the Laboratory Commander identified Five Transformational Imperatives for 2018, outlining his priority areas. One of those priority areas was named Workforce Agility, with specific focus on hiring the workforce AFRL needed to remain technically superior to its adversaries. During this same timeframe, interactions with OPM Director as well as the litany of articles and books published on the gig economy, work place strategies from household-name companies, generations in the workplace, flexible work policies, etc were becoming the norm. Together, these created an environment ripe for research, introspection, curiosity, with an eye towards future application for AFRL to determine not only how the world was changing, but how it would need to pivot in order to stay relevant and technically superior to its adversaries.

From this, the 12 month AFRL Workforce Agility study commenced, ultimately leading to 18 month phased pilots to test new processes identified as best practices by industry or in anticipation of future workforce trends. AFRL is truly experimenting not only with S&T, but with its processes, in order to make data-driven decisions to enable its leaders to hire top tier talent, faster, to collaborate more widely, and to do business differently in this Information Age.

The study occurred in phases, beginning with a parallel literature review process that was predictive in nature, conducted both internally by the AFRL, and externally by Toffler Associates. Results of the literature reviews were strikingly similar, with the findings independently validating one another. Major findings from the dual literature reviews were 6 overarching themes: leveraging foreign national talent while building an internal Science, Technology, Engineering and Mathematics (STEM) talent pipeline; collaboration vs protectionism; whole-person development in S&T; organizing around people and work; embracing continuous learning vs terminal credentials; and meaningful incentives for hiring and continued employee engagement post-hire. A brief summation below of each area expands on these findings.

While small, the ability to identify changing landscapes, best practices, and a willingness to pivot to meet these changes is important not only to AFRL as an organization to stay relevant, but to its current and future employees, to see that leaders are committed to improving processes, changing when needed, and open to feedback and data based decisions on people processes. Gone are the days of making decisions about workforce based on opinion or gut feel, replaced with an era of leaders who seek to be better informed by data, and open to new information from employees or from the outside world and a recognition that agility can only come from the ability to be flexible and change with the world around us.

*Jessica Salyers (The Chicago School of Professional Psychology)*

### **Observability of Work in the Context of Digitalization: Evolution and Management Implications**

Existing studies highlight that digitalization is changing work practices in most developed countries (Colbert, Yee & George; 2016). These studies evidence a number of management issues in the context of digitalization: integration of new organizational modes in the workplace, net generation management, organizational behavior evolution, virtual team management, etc. However, the observability of these changes, in other words, the perceivability manifested by these changes in work and in work-life, and their implications for managerial cognition has not yet been explored.

From a systemic-cognitive approach, this study examines the evolution of the observability of work in the context of digitalization

and the management implications of this evolution through four dimensions of work: the form of work, the content of work, the results of work, and the multiplicity of work-life (Figure 1).

Figure 1: Interactive processes of work and work-life observability (Author, 2019)

**Observability of work form:** In the context of digitalization, work is becoming more intellectual and mental (Quick & Horn, 2008; Price & Hooijberg, 1992). Work activities are progressively shifting from concrete to abstract, comprising different forms of digital objects or interactions. The material tangibility of work has also been declining. Equally, the visual and verbal aspects of work have been declining, as many activities are substituted by written supports that facilitate sharing, diffusing, and achieving the realization of work. Behind a computer screen, the form of work is unified and standardized.

**Observability of work content:** Digital evolution is changing work content, especially through new forms of collaborative organizations. Work is becoming more 'make do' than 'do'. In fact, 'make do' means that work is more easily transferable between professionals, between organizations, or between humans and robots. Activity transfer generates two trends that could appear paradoxical. On the one hand, work is becoming highly standardized through algorithms and automated processes, creating an apparently new category of employees who work on repetitive executions by following standardized processes. On the other hand, collaborative modality provides more opportunities for realizing innovative and creative activities, which might be a source of high-value creation. If these two kinds of work remain indispensable for our society's functioning, they create an important gap in terms of work content. Consequently, the observability of work content is also changing towards two directions. On the one hand, automated and repetitive work is becoming more observable through the computerization of activities and communications. On the other hand, highly qualified or creative work is difficult to observe before its results are realized. Managers need to adapt their observation according to work content. The complexity of observation could be accentuated if these two kinds of work are both performed in the same team. In such cases, managers need to not only vary their observation approach according to work content but also maintain balanced coherence between the two kinds of worker within the team.

**Observability of work results:** Principally due to informational flux, work results are becoming less observable for employees and managers, especially if they work on intermediary processes. Furthermore, work realization is more distributed through a multiplicity of partners and processes, based on computerization and standardization. Employees and managers are, thus, less able to appreciate the value created by their contribution. This might provoke issues regarding motivation and appreciation, which constitute the intrinsic dimension of a job (Danna & Griffin, 1999), potentially impacting their health and well-being (Daniels & Guppy, 1992).

**Observability of work-life:** Work-life becomes multiple and elastic in terms of time as well as in terms of space. It is also becoming more interiorized. Today's individuals anticipate social and organizational pressure and so limit, consciously or unconsciously, the expression of their perceptions of work-life and the negative dimension of their well-being. This self-reframing process might, in some cases, help to surmount temporary difficulties; however, in other cases, it might generate unresolved problems that impact, in the long term, employees' health and well-being.

Through these results, our study explores several potential managerial implications of the evolution of observability of work in the

context of digitalization and highlights the necessity for developing conceptual and empirical frameworks regarding this evolution. The presentation concludes by arguing that organizations should help managers to adapt and develop their observational capabilities by taking this evolution into account.

*Hongxia Peng (Rouen Normandy University)*

### **JobMatchMe—How Psychometrics, Big Data and Psychological Matching Algorithms Help (German) Truck Drivers Find the Job They Love**

In Germany, there are about 500,000 to 1.5m truck drivers and in the European Union, there are about 6m truck drivers in total. As in many countries, logistics companies in Germany are facing two major issues: Firstly, it is difficult for them to find qualified drivers and secondly, drivers change their job frequently (annual fluctuation of 15% to 25%) because they are not satisfied with their job and it is quite easy for them to find a new one.

A startup company, JobMatchMe, founded in 2016, decided to address these issues and to utilize the truck driver market as a proof-of-concept for a highly innovative, disruptive approach to enable low and medium-skilled employees find a job that suits their needs, abilities, and personality.

In the initial stage, the company (then consisting of three founders, a few temps and four business angels) conducted hundreds of interviews with truck drivers (mostly at truck stops along German highways) in order to determine an “Empathy Map” and “Customer Persona”, i.e. to understand truck drivers’ emotional situation, what prompts them to change their job and how they find new ones. Contrary to common understanding, their main issue was lack of appreciation and recognition. Thus, remuneration was not their main reason for seeking a new job, but rather lack of respect and a sense of “not matching with the company”.

On the employers’ side, the situation turned out to be quite dire: On average, they invested almost \$1,000 per month in job ads, usually with meagre response (on average two applications per ad). Turnover and lack of drivers cost them a fortune, as many trucks stand idle instead of moving goods generating turnover.

In 2016, the website “TruckJobs” <http://www.truck-jobs.de> (sorry, in German language only) was launched for drivers and employers. Drivers visiting the website because they are looking for a new job are asked a few simple questions (what are your qualifications/licenses, where and how do you want to work, how much money do you want to earn) and then fill out a brief personality inventory in order to assess four relevant personality dimensions (Conscientiousness, Risk Avoidance, Sociability, Problem Solving).

Employers in turn, describe online job characteristics (what they are looking for, what they offer, what kind of work) and fill out a work culture questionnaire in order to assess the company’s “personality”.

After the driver has entered his personal data, he is immediately presented with a choice of jobs/companies that match his needs and personality (think of: Tinder for truck drivers). In case he “likes” the offer, contact between driver and company is established.

Setting up the diagnostics for drivers and companies turned out to be completely different to the development of psychometrics in an academic context. In detail, several pre-existing scales and items (from the IPIP) had to be revised numerous times in order to determine sufficient psychometric qualities. Luckily, hundreds of drivers visited (and are visiting) the website every week and filled out the questionnaire so the psychometric qualities quickly improved.

As of early 2018, more than 30,000 drivers visit TruckJob’s website each month and more than 3,000 drivers have already found a new job via TruckJobs’ website. TruckJobs has a Facebook website with almost 10,000 followers and more than 1m drivers have watched TruckJobs’ videos on Facebook within a period of three months. Matching drivers and companies is carried out by means of a self-learning algorithm which will be further improved in the future as soon as additional feedback data is integrated into the permanently growing database. Considering truck drivers initially as a mere proof-of-concept, JobMatchMe now strives to expand the concept to other industries (e.g. nurses, hospitality, chefs, hairdressers) to help people find the job they love within 10 minutes.

*Olaf Ringelband (md gesellschaft fuer management-dagnostik)*

### **Automation in Colombia: Assessing Skills Needed for the Future of Work**

Advances in Artificial Intelligence, Machine Learning and Robotics have led to significant interest and discussion about the future of work. Automation has the potential to affect future labor supply, skills demand, current job training, and policy. One recent research approach to understand automation’s effect on the economy estimates the probability of an occupation being automated based on the likelihood that certain occupation-specific tasks are capable of being performed by computers (Nedelkoska & Quintini, 2018; Frey & Osborne, 2017; Arntz, Gregory, & Zierahn, 2016).

Most of the research has focused on occupations in the United States and the OECD countries. Data and research on Latin American countries is scarce. This lack of research hinders proper planning for human capital development and adequate economic policies. The population of Latin American countries is approximately the same as the US and OECD countries combined (WB, 2017). As a result, Latin American countries must address both the commonalities and differences across the labor market through more research.

In this paper we follow the “occupation probability of automation” methodology using data from Colombia. We estimate how many Colombian occupations are vulnerable to being replaced by a robot or machine. We apply a Gaussian Process Classifier on the CNO 2016 (National Occupation Classification), a public database that contains a list of Colombian occupations with a description of abilities, knowledge and tasks.

We identify 21 activities associated with each occupation that computers are unable or incapable of doing in the near future. For example, the activity of “organizing, planning, and prioritizing work” is unlikely to be replaced by a computer. In contrast, the methodology in Frey & Osborne (FO) (2017), identified activities in which computers will soon take over. Our methodology is based on activities in which humans excel, whereas FO’s methodology is based on activities in which computers excel. Our model estimates how the 21 activities are associated with the likelihood of a particular occupation being automated. This output variable is a subjective assessment made by a group of trained scholars on the probability of automation of the Colombian occupations. The output variable is binary, and it represents the occupations likely to be automated (1) as opposed to the occupations in which humans will continue to remain employed (0).

To capture the 21 occupation activities, we match the Colombian database with US Bureau of Labor Statistics Database, O\*NET. Where possible, we match the name and description of the occupations.

CNO is based on O\*Net as a verification for the published version (SENA, 2018), which means that both databases have similar information and can be linked. Despite this connection, there are still some occupations which are unique to either country.

As in Autor (2015), we examine other features that influence how automation affects the country's economy. We posit that the Colombian context is necessarily different to the context in the US or in OECD countries. In a sense, social and cultural factors affect the likelihood of automation and should be included in the analysis and discussion. At the same time, a comparison with the FO results clarifies the differences in Colombian culture, and allows for a discussion of how Colombia should respond to the possible effects of automation.

Preliminary results support the direction of our analyses. We discovered interesting differences between the US and Colombia. For example, agricultural technicians or bartenders were much more likely to be automated in the US labor market than in Colombia. Also, we found that 30.19% of the Colombian occupations are in high risk of automation, 41.02% are in medium risk and 28.77% are in low risk. We show these percentages next to the U-shape probability density function of automation recognized in related studies for the US and European countries. Further analysis will tackle each group of occupations to determine if they require a particular set of skills. With this approach, we will assess the skills and competencies needed to face the automation of jobs more successfully.

This research has several practical implications. Policymakers will be able to better plan for the possible effects of technology and innovation. In particular, we hope they will anticipate new areas of opportunity for skills development and technology investment opportunities. At the same time, universities and other academic institutions should reconsider their education programs and design them in a way that fosters the development of those competencies and skills that will be needed in the future. Managers could develop new careers within their organizations, create new plans for training employees, and identify business opportunities. The results show both similarities and differences in the probability of automation of occupations compared with other countries. The discussion of cultural and political factors enlightens possible reasons for the differences in the results.

Sandra Idrovo (INALDE Business School, University of La Sabana)

## Independence HI



## Psychological Assessment and Treatment Via Technology

### PAPER SESSION

#### Technology as a Sensegiving Agent: A Real-time, Personalized Information System for Emergency Care Patients

**Problem Statement:** Patients in Emergency Departments (EDs) undergo numerous tests and consultations with different professionals and specialists (Pinelle & Gutwin, 2006). This complexity, and the lack of availability and access to information (Israeli Ministry of Health, 2016, 2018) often impedes the patients' ability to make sense of the process (Sensemaking, cf., Weick, 1995). We developed a platform that disseminates personalized, real-time information to

patients about their journey through the ED, and discuss this as a process of technological sensegiving (cf., Gioia & Chittipeddi, 1991) to patients. We show that technology can serve as an effective agent for improving patient sensemaking, satisfaction and trust.

**Procedures:** Having gained a basic understanding of the ED information constraints and requirements, we (developers: researchers and hospital managers) developed an initial design of the information platform (developer sensemaking), which we then used to provide information to patients through a new platform (technology sensegiving), who made sense of their ED journey (patient sensemaking). We collected patient feedback on the initial design (patient sensegiving), and improved and refined the platform accordingly (developer sensemaking and sensegiving). Thus, platform design and development involved an iterative process of sensemaking and sensegiving (cf., Gioia & Chittipeddi, 1991; Weick, 1995) between developers of the platform and ED patients (see Figure 1).

The design and development process was user-centered, comprising moves from initial (concept) design, to pre- and then post-deployment detailed design (cf., Abras et al., 2004; Norman & Draper, 1986). Throughout this process, refinements in sensegiving became more subtle, and sensemaking became more focused, until the platform reached (optimal) convergence, and patients found it satisfying and trustworthy (see Figure 2).

**Analysis and Results: 1: Developer Sensemaking.** Unstructured interviews with managers, medical staff and patients revealed patients' needs for personalized real-time information about their ED journey (Parush et al., 2014). Designing a platform to provide this information required understanding the multiple subsystems and functions (e.g., triage, consultation, MRI, blood tests) that patient ED journeys can comprise. A platform that serves ED patients, needs to predict any potential journey through these subsystems and functions (as discussed next).

**2: Technology Sensegiving.** To use an information platform to give sense to patients, we translated physician-oriented hospital information into terminology that focuses on patient needs and understanding. We identified flows of medical processes (Process mining, cf., van der Aalst, 2011), established detailed routes of patients, and estimated anticipated service duration in each ED station (Queue mining, cf., Senderovich et al., 2015). We developed algorithms that generate updates for each patient based on current ED load (Armony et al., 2015; Batt & Terwiesch, 2015). Algorithmic integration of the process-, load-, and queue data creates updated estimates of durations of patients' medical tests and treatments (Carmeli, et al., 2018), which we show to patients on their personal smartphone (see Figure 3). The platform, which patients access through an SMS, was first deployed in a hospital ED in July 2018.

**3: Patient Sensemaking.** Shortly after deployment, unstructured interviews with ED staff and patients, and direct feedback from ED patients using the platform (as collected over three months), revealed specific aspects of the platform requiring refinement including privacy of content, psychological will to use the platform, and design issues. Between October and December 2018, we distributed surveys to ED patients using the platform. Compared to a control group (patients without platform access; surveys were collected before platform deployment), these patients achieved higher sensemaking, predicting higher patient satisfaction and trust ( $F(16,266)=8.570, R^2=0.340, p<0.001$ ), as well as lower patient aggressive tendencies ( $F(16,270)=2.595, R^2=0.133, p<0.001$ ).

**4: Patient Sensegiving.** Throughout development of the platform,

we relied on patient reactions to the technology, refining our ongoing sensemaking, and the design of the platform. Following patient sensegiving, we continuously adapted various features, to improve the platform usability. For example, the platform now automatically sends a reminder message, and also additional staff, flyers and posters assist in encouraging patient entry and use of the platform.

**Contributions and Practical Implications:** Previous research mainly monitored, and in rare cases, facilitated sensemaking by human agents (Luescher & Lewis, 2008). We show technology can act as an active sensegiving agent, complementing traditional routes of informing, thus reducing the effort required of overstressed medical staff. We show a symbiotic process of simultaneously decreasing refinements in sensegiving and increasing focus in sensemaking as a technology is developed, and illustrate the interdependence of sensemaking and sensegiving of extra-organizational members (patients and developers) and intra-organization members (hospital managers and staff). We also show a highly interdisciplinary nature of developing such a platform; our research integrates Operations Research, Human Factors, Software Engineering, and Organizational Behavior.

The information platform we describe extends current tools that supports multidisciplinary health care to include communication with patients, suggesting that information technology that can alleviate the huge strains on emergency healthcare, and to improve its service delivery (Giebelhausen et al., 2014).

*Anat Rafaeli (Technion - Israel Institute of Technology)*

#### **Schrödinger's Cat Turns to Therapy: Measuring Progress as a Tool for Enhancing Therapy and Standing with the Patient in His Battle Over Consciousness**

Measurements and metrics in psychotherapy is a hot subject for many reasons. The famous of which is the need to evaluate effectiveness of treatment and more commonly cost-effectiveness. Therapy processes that have success and well being criteria are more easily accepted as evidence-based and can more easily be lent for research.

We built an internet-based process-management system named "The Resonator" which follows the patients in their daily life allowing them to incubate the therapeutic content in their consciousness so it can further process and evolve between meetings. Resonators include images, audio and/or videos and supporting text - as well as tools for collecting important information on patients' state.

"Resonate" means that patient and therapist sync and focus their understanding on what the issues are that they are working on, what the common language is for describing these issues and what may help the patients overcome their issue. It also means that therapeutic content stays affective between meeting, fighting for space in patient's consciousness.

Another aspect of The Resonator system, is the bi-directional communication between therapist and patient with the purpose of allowing the patient to register real-time in-vivo life experiences. This kind of communication provides more meaningful therapeutic input for both patient and therapist for evaluating issues in treatment and measuring effectiveness for both the therapist and manager of the clinic—per therapist and on a global basis for the clinic. In turn, the data gathered can be used for research. Patient Resonator-driven reporting provides measurable tools for the patient to define and measure the attributes of their problem, increase awareness to therapeutic content and evaluate more quickly which different therapeutic approaches are more effective.

Last but not least, measuring the therapeutic process is likely to change the process itself. More often than not, the need to decide what we measure will help us realize that we need to redefine what it is that we are treating.

In our paper, we will share the impact of measuring with computer-assisted online systems on the level of "sense of being in treatment" on the patient between meetings and how this increased levels of awareness and connectivity to therapy-vibes impacted the effectiveness of the work.

We will refer to the measuring systems as a "virtual sensor" to a. the level of commitment of the patient to the healing process and resulting from that changes in drop-out levels and to drop-out prevention in the clinic; b. the well being status of the patient. Other factors that we will refer to are: changes in subjective sense of meaning to the therapeutic process in both patient and therapist, focus and sense of direction, adherence to CBT-like assignment, clarity in the therapist mind of the patient's real status, time-to-response in case of changes in the patient's condition, and more. Our experience shows that all these aspects and are influenced by the measurement itself and by the selection of tools of measurement.

*Arnon Rolnick (The Siach Institute for Psychotherapy, Rolnick Center)*

#### **Technology, Bio-Marker and Data Guided Psychotherapy and Efficacy Analytics**

Patients and practitioners alike need valid and reliable information regarding the etiology of mental disorders, diagnoses and effects of treatment that transcend self-report and clinical intuition. Practitioners must be trained to monitor, document and assess mind-body processes that have been found to underlie mental disorders, predict morbidity/mortality trajectories and reflect differential treatment outcome. Failing to recognize the bio-medical consequences of insidious subliminal neuropsychophysiological responses that are associated with mental illness can adversely impact patients. While practitioners may have confidence in their diagnoses and along with their patients equate "feelings" and self-report of symptom amelioration with a return to mental health, advanced biomarker approaches to diagnostics and intervention efficacy testing can call diagnoses and claims of treatment success into question, but, validate them as well. In this applied, practice and research relevant presentation we introduce an integrative, user-friendly, technology supported and biomarker-based exercise protocol to augment psychotherapy. The goal of biomarker based exercise psychotherapy is to 1) document patient autonomic nervous system and cognitive functioning at baseline; 2) during therapy sessions; 3) via patient self-monitoring between sessions and ultimately 4) help determine to what extent exercise facilitates enduring cardiologic, mental and cognitive, biomarker-verifiable change-thresholds that are associated with angiogenesis and neurogenesis (processes that are central to integrative-health). The aforementioned exercise protocol's procedures, instrumentation/technology and analytics will be discussed and demonstrated. We will also introduce practitioners to our Universal Clinical Trial (UCT), an applied and research protocol for collecting "big data" on psychotherapy treatment outcomes (efficacy) in the context of technology derived biomarkers. Our presentation/talk is intended to convey how technology can be used by psychologists and allied mental health practitioners to better assess the cardiologic, mental and cognitive status of patients at intake, the impact of psychotherapy on mind-body processes during and outside of in-office

sessions, longitudinally. We are not presenting study data per se, but will discuss clinical case-study findings that were derived using our protocol. The protocol's steps are as follows: 1) heart rate variability (HRV) baseline, stress and neurocognitive testing at intake, paying special attention to the SDNN measure (the HRV Index & validated biomarker correlate of cardiologic, mental & cognitive status) 2) continuous monitoring of HRV for the duration of a psychotherapeutic session 3) patient self-monitoring of HRV at home and in the context of prescribed exercise therapy and 4) intervention outcome (psychotherapy & exercise efficacy analytics at 3,6,9 & 12 month endpoints). The clinical goal is to raise SDNN, a millisecond (ms) reading by 3-5 ms. a month in patients whose baseline SDNN is < 30ms. (a red-flag finding and independent cardiologic risk-factor & correlate of mental illness) and increase cognitive performance. Increases in SDNN to a high age-normative level through aerobic-exercise is expected to lead to a reduction in psychological symptoms (e.g., depression) as well as the angiogenesis (enhanced cardiac functioning) and neurogenesis that is associated with "brain fitness," cognitive protective and enhancing mechanisms. We are in the process of organizing a so-called "Universal Clinical Trial" (UCT) that is designed to attempt to validate our very promising protocol that has demonstrated, clinically anecdotally, that SDNN is a highly predictive biomarker correlate in the mental health/health and cognitive performance (cognitive status/brain health) equation. We would like to introduce our UCT through this conference and encourage psychologists and allied practitioners to participate and become trained in technology-based and biomarker-guided approaches to psychotherapy in the interest of advancing "best practices." In this applied, practice and research relevant presentation we introduce an integrative, user-friendly, technology supported and biomarker-based exercise protocol to augment psychotherapy. The goal of biomarker based exercise psychotherapy is to 1) document patient autonomic nervous system and cognitive functioning at baseline; 2) during therapy sessions; 3) via patient self-monitoring between sessions and ultimately 4) help determine to what extent exercise facilitates enduring cardiologic, mental and cognitive, biomarker-verifiable change-thresholds that are associated with angio-and neurogenesis (processes that are central to integrative-health).

*Roland Carlstedt (McLean Hospital/Harvard University)*

---

## Lafayette Park

### Physiology and Psychology

#### PAPER SESSION

##### **Wearable Physiological Sensors: Usability and Privacy Factors That May Affect Their Adoption by Future Soldiers**

Wearable sensors can enable an individual's current physical status and well-being to be monitored (Majumder, Mondal & Deen, 2017). The Army has a vested interest in the well-being of soldiers, and the use of wearable sensors by soldiers in small combat units might enable Army leaders to remotely monitor physiological signals to detect impending issues such as heat strain (Tharion et al., 2013; Scataglini, Andreoni & Gallant, 2015). An objective of the current

research was to better understand factors relevant to sensor usability and comfort, with an eye to identifying the factors that would best predict whether future soldiers would readily adopt the use of wearable sensors (of different forms) and comply with instructions to wear them in the field.

The sample for this study was drawn from the population of university-student Cadets at West Point (the United States Military Academy), who will shortly become Army Officers. There are a variety of wearable sensors in use by West Point Cadets—for example, some were purchased by Cadets for their personal use, some were provided by coaches to the members of the various Varsity athletic teams, and some sensors were worn by volunteers in research studies. In an on-line survey, Cadets who had experience with a wearable sensor were asked to rate, via 7-point Likert scales, various usability aspects of their wearable device, including: the devices' comfort; fit; ease of putting it on/taking it off; degree of interference with various activities; and how willing they might be to wear such a device while sleeping, and/or continuously for multi-day periods, and/or in field and combat settings. Users were also asked about the degree to which having access to their own data would increase their willingness to wear a sensor, and if so, what type of data metrics they'd be most interested in. They were also asked how they felt about peers and/or hackers having potential access to their data, and leaders using their data for purposes other than monitoring their well-being during a mission (e.g., fitness and performance evaluation).

The variety of sensors in use at West Point enables the comparison of different form factors (e.g., wrist-worn sensors, foot pods, pods/pucks in clothing such as halter tops, etc.). The analyses include a between-groups ANOVA to compare these different form factors for each of the usability questions mentioned above. The results clearly will have practical implications in terms of providing insight about the willingness of future soldiers to adopt and comply with the use of such sensors, and whether particular form factors might be more readily adopted. Particularly relevant questions for adoption/compliance included those about whether users would be willing to commit to long term (multi-day) wear, whether they had concerns about task interference in field settings, and whether they had specific desires or concerns about data access. Recommendations about the most readily-adoptable sensor format and data protection/use policy will be made based on these results.

Data collection is currently on-going but collection and analysis will be complete by May 2, 2019 (well in advance of the APA:TMS conference), because this research project is part of a Senior-Year Graduation requirement for the student authors.

*Aryn Pyke (U.S. Military Academy at West Point)*

##### **Good Vibrations: Consumer Responses to Technology-Mediated Haptic Feedback**

Manufacturers often incorporate haptic feedback (tactile technology that applies forces, vibrations, or motions to the skin) in consumer gadgets (Bark et al. 2008). However, despite the prevalence of such device-delivered haptic feedback, very little research has examined consumer responses to it. We argue that it is important to investigate what additional psychological and behavioral consequences might stem from these sensations. Some scholars in computer science suggest that device-delivered haptic feedback can symbolize a human's touch under very specific conditions (e.g., if users are explicitly told that the sensations represent the touch of another person, Haans

and IJsselstein 2006), and social psychology research shows that incidental interpersonal touch can shape people's behavior and judgments in various ways (Gallace and Spence 2010). We suggest that in addition to simply alerting consumers, haptic feedback accompanying communications may also play an additional role: generating a sense of "social presence" (Qiu and Benbasat 2009) in what may otherwise feel like a cold technological exchange. This sense of social presence makes the communication itself more meaningful, and as a consequence, more effective in motivating behavioral responses.

We focus our exploration on one important area of consumer performance: physical fitness. Physical fitness is an externally-relevant context to investigate device-mediated communications given the skyrocketing adoption of health and fitness apps and wearable fitness trackers in the marketplace (Lamkin 2016). In addition, because we expect the potential positive effects of haptic alerts on performance to be driven by an increased sense of social presence, it was important to select a context in which social support reliably improves performance. A well-established literature has demonstrated the positive effects of social support on physical performance and exercise (for a meta-analysis, see Carron, Hausenblas and Mack 1996). Given that people show increased motivation and performance on physical fitness activities when in the physical presence of social support, it is compelling to explore whether social presence activated through technology-mediated incidental touch might also improve attitudes and increase voluntary compliance in this consequential domain.

We tested our predictions across three laboratory studies. In study 1, participants (N=60) received mobile phones set to either beep or vibrate upon receiving text messages. Participants received a series of text messages intended to provide encouragement (e.g. "You're doing great! Keep it up.") while attempting a physical challenge (balancing on one leg for 5 minutes). Binary logistic regression confirmed a significant main effect of message alert on performance ( $\chi^2=4.29, p<.04$ ), in that participants in the vibrotactile alert condition were more likely to successfully remain on one leg for the duration of the challenge than those in the auditory alert condition.

In study 2 (N=86), we explored the impact of vibrotactile alerts in a different product category (smartwatches) and a different task (a "steps challenge," where participants marched to get as many steps as possible in five minutes). Participants again received a series of encouraging text messages. Messages were accompanied by one of three alerts: auditory vs. vibrotactile vs. auditory+vibrotactile. Adding the third condition allowed us to test whether Study 1 effects were driven by the inclusion of vibrotactile sensation or by the absence of auditory cues. Pedometers on participants' smartwatches recorded the number of steps achieved, and ANOVA results confirmed a significant main effect of message alert on the steps achieved by participants ( $p<.02$ ). Planned contrasts demonstrated that while there was no difference in performance between the vibrotactile vs. auditory+vibrotactile conditions, subjects in both these conditions outperformed those in the auditory condition, suggesting that improved performance was driven by the inclusion of the haptic cue.

Study 3 (N=56) was similar to study 2, but in addition to measuring performance, we collected several potential process measures: social presence ("Who do you think was sending the text messages?" rated on whether attribution was to an automated system/machine or a human source), message evaluation (e.g. "The text messages were nice,"). Further, we measured technological self-efficacy as a potential moderator, since previous research suggests low

self-efficacy should exaggerate the value of haptic feedback. ANOVA results confirmed a significant positive effect of vibrotactile alerts on the number of steps participants performed ( $p<.02$ ). While there was a marginal effect of vibrotactile alerts on arousal ( $p<.07$ ), this did not mediate the effect on performance, and the impact of vibrotactile alerts on performance was still significant when controlling for arousal ( $p<.04$ ). A sequential mediation analysis (Hayes 2013, model 6) supported our hypothesized path: Message Alert x Technological Self-Efficacy interaction  $\rightarrow$  Social Presence  $\rightarrow$  Message Evaluation  $\rightarrow$  Task Performance (95% CI excluding zero). This suggests that especially for those low in technological competence, the positive effect of the vibrotactile alert on task performance was due to increased feelings of social presence and improved evaluation of the communication respectively.

In sum, our studies demonstrate that haptic alerts can increase feelings of social presence and improve performance on related physical tasks.

*Ana Valenzuela (Baruch College, CUNY)*

### **Evaluating Human Factors in the use of Lower Limbs Exoskeletons: A contribution to the first European benchmarking framework**

Lower-Limbs Exoskeletons (LLEs) are already used in different fields of application (i.e. industrial, medical and military) and more advanced models are currently under development. In healthcare, LLEs are used for both rehabilitation and assistive purposes with the aim of stimulating motor recovery and enabling participation in activities of daily living (Molteni et al., 2018). Using a morphological classification of robots (Moulières-Seban, Bitonneau, Salotti, Thibault & Claverie, 2017), exoskeletons may be considered as a "robot" worn by a human to improve its performance or mitigate his handicap, thus human-robot collaboration aspects are of crucial importance.

In their meta-analysis, Miller, Zimmermann and Herbert (2016) compared studies regarding the use of LLEs with spinal cord injury patients, concluding that the use of LLEs allows safe ambulation in real-world settings. However, studies under consideration mainly focused on training protocols, ambulatory performance and physiological demands, while hardly considering Human Factors in the interaction with the system. Furthermore, Torricelli et al. (2015) highlighted that there is a lack of understanding of the users' experience in the evaluation of bipedal locomotion for rehabilitation and assistive purposes. To our knowledge, human-factors aspects in the interaction between users and medical LLEs are not sufficiently investigated. Additionally, there is a lack of standardized protocols to evaluate efficacy and performances of LLEs needed to evaluate specific characteristics (e.g. dependability, adaptability, configurability, etc.) and abilities (e.g. interaction, perception, decisional, etc.) thus allowing to benchmark different models. Benchmarking LLEs allows to verify how different technologies and design choices may affect system abilities, user experience, usability and, consequently, systems' performances.

The present study proposes the development of a Human Factors benchmarking framework for medical LLEs as part of the STEPbySTEP project (a sub-project of EUROBENCH, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No.779963). STEPbySTEP will develop a modular and sensorized staircase testbed to benchmark LLEs abilities, performances and the effects on user experience during ascending/

descending stairs. The current paper will specifically focus on the development of a standard procedure to assess medical LLEs' usability and their impact on users' cognitive workload and trust.

Most of the currently available medical LLEs are designed without taking into consideration human-factors and human-technology interactions aspects. LLEs should require minimal user input, while fostering effective, efficient and satisfactory interactions with the user (Stirling, Siu, Jones & Duda, 2018). In particular, there is currently no framework in literature that allows for an evaluation of medical LLEs usability and the impact on users' cognitive workload and trust. The ascending/descending stairs use case selected in the STEPbySTEP project is particularly relevant due to the frequency it is encountered in everyday context and for the challenge that represents for physically impaired people.

Different technologies, interfaces and design characteristics of an LLE may affect usability, trust and cognitive workload of the user (Stirling et al., 2018). The Human Factors assessment framework developed in STEPbySTEP aims to provide a standard method to measure the performance of an LLE in terms of operational benefits, such as the capability to collaborate with the human, as well as measuring users' trust, cognitive workload and system usability. Baseline data will be gathered from preliminary tests performed on one or a small group of patients (level of lesion D4-L1) walking with an exoskeleton (ReWalk™) and climbing stairs.

The following data will be collected: users' acceptability of the system (Van der Laan, Heino & De Waard, 1997); users' experience; users' requirements and perceived safety; users' trust towards the system (Charalambous, Fletcher & Webb, 2016), cognitive workload (Hart & Staveland, 1988); observation of behavioral indicators (i.e. time spent, goal achievement, movement accuracy, number of body movements); System usability (Lewis & Sauro, 2017); Comfort and pain; Interface design analysis.

Metrics will be assessed through a software with a user interface that allows evaluators and testers/participant to input data. Scores will be calculated by the software and integrated with physical ergonomics and biomechanical metrics to extract the overall benchmarking score.

The present benchmarking framework could have a number of practical implications. First of all, it will allow to systematically distinguish between well-designed exoskeletons with features that allows to improve cognitive performance, helping the users to shift cognitive resources dedicated to physical effort to the actual activity (Bridger et al., 2018), user's trust and usability. Furthermore, the framework could help LLEs developers to adopt a User Centric Design approach to test prototypes and gather useful insights already during the design phase.

*Federico Fraboni (Department of Psychology, University of Bologna)*

### **Improving Psychological Resilience with Cognitive Retraining Methods using EEG brain network biomarkers: Examples from NASA Lunar/Martian Habitat, Clinical Mood and Substance Abuse Studies**

Paper Focus: Opportunities to increase coping capabilities at the neurophysiological level and the resulting impact on psychological resilience are appealing and potentially transformative, both clinically and in everyday life. The subject of this paper is focused on reporting the results of one such method that we, the authors have employed during three UND/NASA sponsored crewed Lunar/

Martian Analog Habitat simulator missions, as well as separately in clinical populations that include depressed and/or substance abuse subjects (n=200).

The method employs the use of the most current Cognitive Rehabilitation Training (CRT) methods that incorporates a Brain Computer Interface (BCI) which provides neural network performance integrity metrics (nPIMs). These metrics are derived from the neural network system that supports the cognitive function being trained. The CRT methodology is implemented as a set of computer activities to engage the desired training cognitive functions based on classic neuroscience experiments found in the literature. The BCI interface informs the trainer, the user, and CRT activity in real time, regarding current neural network performance integrity status based on the user's present nPIMs state.

The CRT activity incorporates a performance leveling algorithm (PLA) to adjust the intensity of the activity by rendering the pursuit to be either more or less intense. Unique in our method is that the PLA encompasses both nPIMs and behavioral responses (response times, accuracy) to adjust the level of intensity play of the activity. This is based upon the current real-time performance ability of the user that is required to properly engage the long term potentiation (LTP) and long term depreciation (LTD) network learning rules [1-7]. The intention of the performance leveling algorithm adjustment is to adjust the level of activity play to a comfortable level, thus allowing the user to progress through the activity successfully while at the same time focusing on developing and/or strengthening the performance integrity of the neural system being trained.

Why-Coping Skills - Background: Due to ever-changing environmental demands, the human brain has evolved to fine-tune our behavior under variable environmental conditions to help us generate appropriate responses essential for task performance, goal attainment, social interaction, and avoiding danger. However, when too much stress and/or adversity enters one's life, one's ability to respond in a resilient manner is often impacted. Psychological resilience is a construct that represents an individual's ability to positively adapt and respond to stress and adversity while maintaining proper mental health and well-being [8]. Moreover, psychological resilience has been linked to how one copes with adversities and coping style [8-12].

Coping skills are a set of mental and behavioral strategies implemented when facing stress or traumatic experiences. Various studies have uncovered meaningful psychological resilience relationships. Psychological resilience has been shown to be connected with the psychological factors that involve positive and negative affect, emotional regulation and flexibility, self-esteem, inhibitory control, and more [13-16]. Related to these studies, avoidance coping styles have been linked to a disposition to foster psychiatric disorders such as PTSD, anxiety, and major depression, while problem-oriented coping skills are correlated positively with well-being and a higher quality of life. Coping styles constitute an important determinant of resilience.

Training Targets: From a neuroscience perspective, Santarnecchi et al. (2018) investigated coping styles based on the brain's neural network functional connectivity underpinnings. The authors correlated psychometric scores of coping with resting-state fMRI data in 102 healthy adult participants. Significant links between the propensity to adopt different coping styles and specific network functional connectivity profiles for regions belonging to the default mode (DMN) and anterior salience (AS) networks were found. Additional findings included reduced negative correlations between AS and

DMN nodes. This particular finding was considered to explain the variability in the coping style that is related to avoiding problems while focusing on the emotional component of the stressor at hand, in place of trusting on cognitive resources.

Other authors have further found psychological resilience to be positively correlated with the connectivity strengths of the left insula and the right parahippocampus, which are both involved in the self-evaluation processes [13-24]. Equally important, psychological resilience has also been shown to be positively correlated with the strength of the left orbitofrontal gyrus (OFC) and the left inferior frontal gyrus (IFG) connectivity, associated with the adaptable use of emotional resources and flexible control when processing affective information [13-24]. Likewise, resilience has also been shown to be negatively correlated with the strength of the left OFC and the right precuneus connectivity, shown to be involved in rumination of negative self-related thoughts [13-24]. XXX et al. has shown that, the left OFC-IFG connectivity mediates positive resilience effects in support of the view that positive affect assists resilience by broadening one's attention and promoting flexible thinking and coping abilities [13-24].

*Curtis Cripe (Fordham University)*

## Independence E

### Algorithms and Advice

#### PAPER SESSION

#### Base Rates, Outcomes, and Information Representation in a Bayesian Inference Task

Many contemporary digital applications require people to estimate the likelihood of outcomes using Bayesian inference. For example, an online patient portal might report the results of a patient's test for a disease and the rate of disease occurrence in the population, leaving patients to determine their likelihood of having the disease. Much research has shown that people find such tasks difficult (Tversky, 1982). Understanding those factors that influence performance could facilitate the development of digital applications that better support Bayesian inference.

One strategy to improve Bayesian performance has shown promise: changing the information representation of the task. For example, changing the representation from probabilities to numerical frequencies can dramatically improve accuracy (Barbey & Sloman, 2007; Gigerenzer & Hoffrage, 1995). Presenting information graphically in static and/or interactive forms, sometimes but not always improves performance, compared to non-graphical forms (Böcherer-Linder & Eichler, 2017; Garcia-Retamero & Cokely, 2011; Khan, Breslav, Glueck, & Hornbæk, 2015; Micallef, Dragicevic, & Fekete, 2012; Ottley et al., 2016; Sloman, Over, Slovak, & Stibel, 2003; Starns, Cohen, Bosco, & Hirst, 2018; Tsai, Miller, & Kirlik, 2011). Thus, in terms of building digital Bayesian experiences, it is not clear that there is one best information representation.

It is known within the human-computer interaction literature that interface features, and task characteristics together influence users' mental representations of problem spaces, their choices of problem solving strategies and ultimately, their levels of performance (Norman, 1986; Poor et al., 2016). Within the digital Bayesian application realm, we hypothesize that information representations

(interface features) and problem characteristics (task features) might combine together to influence performance. Our recent work on a Bayes-related task demonstrated that moving from a non-graphical to graphical representation enhanced people's ability to discern between inferences that were consistent with Bayes' rule and those that were Bayes-inconsistent (Anderson et al., 2019).

Here, we present a study of Bayesian inference in which we varied the information representation, the base rates, and correct outcomes. We used the problem described in Figure 1. The levels of the Representation variable, which varied between participants, were Probability, Numerical Frequency and Graphical Frequency, shown in Figures 2, 3, 4, respectively. The levels of Base Rate ("A > B", "A = B" and "B > A") varied within participants, as did the levels of Correct Outcome. The correct outcomes were, "Given that a chip is hot:" (i) "It is likely a Brand A chip", (ii) "Brands A and B are Equally likely," and (iii) "It is likely a Brand B chip."

We collected our data using Qualtrics, similar to Khan et al. (2015); Micallef et al. (2012); Ottley et al. (2016). Participants answered nine Bayesian inference problems produced by crossing the Base Rate factor with the Correct Outcome factor. The problems were randomly ordered for each participant. Each participant was randomly assigned to a Representation condition. We only accepted complete submissions, submitted on a computer (rather than a cell phone or tablet), leaving 27 Probability, 24 Numerical Frequency and 24 Graphical Frequency participants. Participants were drawn from undergraduate classes at a local university. The local institutional review board approved the research; participants were treated in accordance with the Declaration of Helsinki.

The overall non-weighted accuracy rate was 57.49% (SD = 29.82%). We multiplied accuracy (coded -1 or 1) by confidence (an integer ranging from 0 to 4) to produce Confidence-Weighted Accuracy, which could range from -4 to 4. We analyzed our data using a repeated measures ANOVA: Confidence-Weighted Accuracy was the dependent variable, Base Rate and Correct Outcome were within-participant factors, and Representation was a between-participant factor. The main effect of Representation was not significant. The three-way interaction of Base Rate, Correct Outcome and Representation was significant (Greenhouse-Geisser  $F(6.659, 239.747) = 5.626, p < 0.001, \text{partial } \eta^2 = 0.135$ ) (See Figure 5). Our interpretation of the interaction is that the stability of the means across all problem conditions was greater for the Graphical Frequency condition (yielding no pairwise differences) than in the non-graphical conditions. The non-graphical conditions yielded a complex pattern of means. For example, when the correct outcome was A = B (i.e., "Brands A and B are Equally likely"), the means for the non-graphical conditions were either above or below those for the graphical condition, depending on whether the base rate condition was "A = B," "A > B," or "A < B" (see Figure 5).

In summary, we found that performance depended not just on the representation, but on a combination of representation and problem conditions. That graphical representation yielded little variation in performance across problem conditions, suggests that such representation supported the same strategy across problem conditions. Further work will investigate the details of such a strategy, and how the strategy could be effectively integrated into digital Bayesian inference applications.

*Laura Leventhal (The University of Findlay)*

### Algorithm Does Not Lead to Prevalent Information Avoidance

Statement of the problem: In a series of studies, we will investigate whether people are more likely to avoid information when that information comes from an automated system compared to a human agent. We think psychological motives such as privacy, vulnerability, accuracy and effort may make people less likely to avoid information from machines. In contrast, issues of trust may make it less likely to avoid information from humans.

Procedures, Analyses & Results: All of our studies are scenario based. In Study 1, we constructed our scenario from the classic “ostrich effect” (Karlsson et al., 2005) and found participants’ choice to avoid financial information is statistically insignificant between the human and algorithm condition. There were two conditions in a between-subjects design. We modified our phrasing, tightened our design, and replicated the study (Studies 2 & 3). There were four conditions in a 2 (stocks up, stocks down) x 2 (machine, human) between-subjects design. We used a two by two ANOVA analysis to test the main effect of stock market gain/loss, main effect of type of phone line, and effect for the interaction of the two on the number of people choosing to call the phone line. However, the results still refuted our initial hypothesis. This null effect persisted when we changed the scenario to health information (Woolley & Risen, 2018) in Study 4 where we used two conditions in a between-subjects design: algorithm- participants learnt they can use a tablet computer about the calories human—participants learnt they can ask a restaurant worker about the calories. Our primary analysis was a chi-square test which compared the percent of participants choosing yes in each condition. Study 5 featured an AI and human news anchor broadcasting 6 different news reports. There was one condition in a within-subjects design: participants learnt they can either watch the news from a human presenter or a robot presenter. The 6 news items were presented in randomized order. A linear regression predicting participants’ preferences with a dummy representing each condition, no intercept, and standard errors clustered by a participant ID differences indicated preferences vary significantly between the human versus algorithm condition.

Practical Implications & Conclusions: We interpret our findings as indicating a null effect of using machines to provoke more information avoidance. On the other hand, just because we didn’t find anything in the first four studies doesn’t mean people have the same preferences on where they retrieve certain information. In fact, people do care where they obtain information, even if those preferences don’t necessarily map into the calculus of information avoidance. Future research can dive deeper into analyzing why those preferences exist and ways in which they will influence people’s desire to seek information. Our study is relevant to current debates on artificial intelligence. The proliferation of AI in banking industries has generated speculation on ways in which future financial information will be delivered. Since information serves as a crucial part of human decision making, understanding the psychological mechanisms that govern its functions can help people make wiser judgments. Another practical implication is analyzing students’ ability to retain information—a key contributor to academic performance—when robots dominate the educational curriculum. Giant companies such as Amazon and Alibaba can use our findings to decide whether AI-oriented commercialization will generate more profits.

Relevant literature to establish the research context of the current proposal: People frequently avoid information in their daily lives.

Around 12% to 55% of people who undergo testing for HIV fail to return to learn whether they were infected (Hightow et al., 2003). Furthermore, many people choose not to learn their risk for cancer when provided an opportunity to do so (Keogh et al., 2004; Ropka, Wenzel, Phillips, Siadaty, & Philbrick, 2006). In the financial world, investors log into their accounts more frequently on days when the market has gone up but remain inattentive when they suspect unfavorable news (Karlsson, Loewenstein, & Seppi, 2009). Psychologists have identified several motives that contribute to this phenomenon—protecting intuitive preferences (Woolley & Risen, 2017), diminishing unpleasant emotions (Boen, Vanbeselaere, & Feys, 2002) and reducing self-threat (Frey, Stahlberg, & Fries, 1986). Recent research also highlights that people treat machines and humans asymmetrically while making forecasts (Dietvorst, Simmons, & Massey, 2015). Drawing on this literature, we will investigate whether people are more likely to avoid information when that information comes from an automated system compared to a human agent. This will connect current debates in psychology with the rise of artificial intelligence.

*Wenjie Han (The University of Chicago)*

### Algorithm Appreciation: People Prefer Algorithmic to Human Judgment

Although people often receive advice from other people, the rise of “big data” has increased both the availability and utility of a new source of advice: algorithms. The superior accuracy of algorithmic judgment relative to human judgment (Dawes, Faust, & Meehl, 1989) has led countless organizations to invest in the power of algorithms—scripts for mathematical calculations—to sift through data and produce insights. This widespread reliance on algorithmic advice appears at odds with work suggesting aversion to algorithms (Dawes, 1979). Nevertheless, the idea of aversion to algorithms is prevalent and many articles advise business leaders on how to overcome it (Harrell, 2016).

Counter to this notion, results from eight experiments suggest that lay people show “algorithm appreciation.” Specifically, people adhere more to advice when they think it comes from an algorithm than a person when making estimates and forecasts. In our experiments, we used the judge-advisor system paradigm (Snizek & Buckley, 1995), which allows us to benchmark people’s reliance on advice from an algorithmic advisor against their reliance on advice from a human advisor. This paradigm importantly allows us to take into account people’s general tendency to discount advice.

People showed “algorithm appreciation” when making estimates about a visual stimulus (Experiment 1A; N = 202), forecasting the popularity of songs (Experiment 1B; N = 215) and forecasting romantic matches (Experiment 1C; N = 286). We obtained this effect across multiple operationalizations of “human” vs. “algorithmic” advice. Numeracy appears to be a mechanism for reliance on algorithms; those in Experiment 1A who scored higher on a math test (Schwartz, Woloshin, Black, & Welch, 1997) relied more on algorithmic advice.

Yet, when we provided Judgment and Decision Making researchers from the society’s mailing list with the survey materials from Experiment 1C (Experiment 1D; N = 119), they predicted the opposite results. Although mTurk participants displayed algorithm appreciation when predicting romantic attraction, researchers expected them to display algorithm aversion. The idea that people are averse to algorithmic advice is evidently pervasive, even among expert researchers.

Reliance on algorithms was robust to presenting the advisors jointly or separately (Experiment 2; N = 154). However, algorithm

appreciation waned when people chose between an algorithm's estimate and their own (versus an external advisor's; Experiment 3;  $N = 403$ ). These tests are important because participants could improve their accuracy by relying on algorithms more than they already do (Experiment 4;  $N = 671$ ) relative to a normative standard.

Experiment 5 examined how decision makers' own expertise influenced reliance on algorithms. National security professionals who make forecasts on a regular basis responded differently to algorithmic advice than lay people did. Experts ( $N = 70$ ) relied less on algorithmic advice compared to lay people ( $N = 301$ ), heavily discounting any advice they received. Although providing advice from algorithms may increase adherence to advice for non-experts, it seems that algorithmic advice falls on deaf expert ears.

These results suggest that the story of algorithm aversion is not as straightforward as many people, including current scholars, may think. Providing algorithmic advice appears to be one way to increase lay people's adherence to advice, which is useful considering that people tend to discount advice from others (Bonaccio & Dalal, 2006).

*Jennifer Logg (Harvard University)*

## Independence FG



## Virtual Reality for Health and Well-Being

### PAPER SESSION

#### Virtual Reality Can Increase Empathy: A Preregistered Meta-Analysis.

Virtual Reality (VR) has been touted as a simple intervention to increase empathy. VR replaces our own sensory environment with a new perspective, not only allowing us to walk in another's shoes, but climb inside their skin. In this way, VR may help people to empathize with others in circumstances they find it difficult to imagine. Although popular, evidence on its efficacy is mixed. This systematic meta-analysis aimed to collate evidence across disciplines to estimate the effect of VR on empathy and investigate moderators of this effect size.

5583 articles were initially identified from keyword database searches, 4929 of which were excluded as irrelevant based on screening of titles and abstracts. Of the remaining articles, 37 were found to meet all the inclusion criteria. Broad inclusion criteria were used allowing any type of VR and any dependent measure of empathy. Empathy was defined to include understanding the mental states of others and/or having an emotional reaction to the mental states of others. VR must use computer technology to augment one or several of the human senses in order to immerse the participant in another's experience.

A random effect meta-analysis was conducted based on data from 43 studies (some articles reporting multiple studies). The standardized difference in means was 0.43 [CI 0.29, 0.57] ( $z = 5.91$ ,  $p < 0.001$ ) indicating that VR has a significant positive impact on empathy. The 95% prediction interval is -0.37 to +1.23, suggesting that the true effect size in 95% of cases will fall somewhere in this range. The mean effect size is moderately impressive, so on average empathy is improved by an amount that may have a substantive impact. However, the dispersion of effects about this mean is substantial and this variation exceeds that which would be expected by random

variation ( $Q(42) = 301.69$ ,  $p < 0.001$ ;  $I^2 = 86.08$ ). Subgroup analyses were used to explore this variation.

These analyses revealed that auditory simulations and head-mounted simulations both significantly improved empathy (StdMAud = 0.88,  $p < 0.001$ ; StdMHMD = 0.43,  $p < 0.001$ ) whereas desktop VR did not (StdMDesk = 0.21,  $p = 0.132$ ). VR significantly improved affective empathy (StdMAff = 0.44,  $p < 0.001$ ) but not cognitive empathy (StdMCog = 0.19,  $p = 0.064$ ). In addition, VR significantly improved empathy for healthy adult and clinical populations (StdMAdu = 0.42,  $p < 0.001$ ; StdMClin = 0.96,  $p = 0.008$ ) but not for children (StdMChi = 0.34,  $p = 0.088$ ).

Finally, it is important to consider the utility of VR in comparison to cheaper alternatives. VR is more effective than no treatment (StdM = 0.80,  $p < 0.001$ ), video (StdM = 0.36,  $p = 0.045$ ) and perspective-taking instructions (StdM = 0.31,  $p = 0.030$ ) but is not significantly more effective when compared to reading about others (StdM = 0.31,  $p = 0.056$ ) or witnessing them in real life (StdM = -0.25,  $p = 0.38$ ).

On the basis of this meta-analysis, we conclude that VR experiences can increase empathy, and that such interventions may have a substantive impact. These findings support the use of VR in a variety of situations and with a variety of populations. However, these results also suggest several important boundary conditions to this effect.

Certain types of VR appear much more effective than others at improving empathy. Considering the expense of head-mounted units, many interventions have been designed with Desktop VR that runs on a normal computer monitor. These results indicate, however, that desktop VR does not lead to increased empathy. Interestingly, however, auditory only simulations not only led to an increase in empathy but had some of the largest effect sizes suggesting that low-tech solutions may be a viable alternative. Similarly, other empathy interventions such as reading about stigmatized groups or witnessing them in person were found to be as equally effective as VR.

VR appears effective only at increasing affective empathy, not cognitive empathy. The rush of emotions elicited on behalf of another does not appear to translate into a greater ability to understand others or take their perspective. This may be due to the very nature of VR which immerses the user in a foreign environment providing clear, overt sensory information about the experience without requiring the user to mentally recreate another's experiences. Virtual reality may relieve the burdensome cognitive requirements of traditional perspective-taking, yet, perhaps it is this burdensome engagement that is the key to improving cognitive empathy.

We hope this meta-analysis and initial evidence of crucial boundary conditions will help charities and businesses considering using VR make informed choices and will assist VR developers to further refine this technology to ensure that it maximizes its prosocial potential.

*Alison Jane Martingano (The New School for Social Research)*

#### VR Game Play in a World of Ads: Is Mindfulness a Cause or an Effect?

Video games provide leisure and rest for many people (Abhishek & Suma, 2014). Two-dimensional video games are slowly giving way to Augmented and Virtual Reality (VR), digital immersive gaming experiences (Chittaro & Buttussi, 2015; Akchelov & Galanina, 2016). Many teenage and young adults prefer video games to television (Bernhard, Zhang, & Wimmer, 2011). Since this group make up the

majority of the population in many regions, advertisers have had to turn their attention to games for product and service promotion (Terlutter & Capella, 2013). In-game ad revenues have been below the projected figures (Herrewijn & Poels, 2013). One reason for the slow growth and scepticism is that advertisers do not have accurate measures of ad effectiveness (Bernhard, Zhang, & Wimmer, 2011). For gamers, there is a cost to the task switching involved in maintaining attention on the game, the game controls, and the ads (Bernhard, Zhang, & Wimmer, 2011).

This study explores how advertisers can optimize ads (to achieve greater awareness and increased intention to purchase), without annoying or alienating gamers. It explores whether the Mindfulness of gamers affects their ability to focus on the game and the ad, without a negative impact on gaming outcome. We will also examine whether stress results from gameplay and the impact of VR ads and Mindfulness on the resultant stress.

Player Expertise, Performance, Stress, Ad Awareness and Intention to Purchase. All else being constant, game performance will depend on whether a player is an expert or a novice. Expert players would score higher in-game performance, and novice players will have lower scores (Hypothesis 1).

Playing video games requires focus and attention (Herrewijn & Poels, 2013; Gangadharbatla, Bradley, & Wise, 2013). The competitive nature of many games, and the effort to make optimal use of scarce resources put a mental strain on players. This strain may manifest as increased pulse rate (Jeong, Bohil, & Biocca, 2011; Hasan, Bègue, & Bushman, 2013). The immersive nature of VR games may make players feel threatened by the 'dangers' therein (Grigorovici & Constantin, 2004; Hasan, Bègue, & Bushman, 2013). This can cause distress depending on whether the player is an expert or a novice (Hypothesis 2).

According to the load theory of attention (Lavie, Hirst, De Fockert, & Viding, 2004), there is a limit to task-switching. When a player notices an ad during gameplay, an attempt would first be made to balance out the increased mental load (Grace & Coyle, 2011). When this fails, the ad will be likely ignored (Hypothesis 3).

The presence of Ads might increase the cognitive and mental demands hence leading to more stress (Hypothesis 4).

Moderating Effect of Mindfulness. Mindfulness is the non-judgmental awareness and attention to the present moment (Brown & Ryan, 2003). A mindful gamer will be able to acknowledge an ad, without anxiety and judgement, and immediately and easily return attention to the primary task of gaming. This transition should be so seamless that the presence of the ad does not negatively impact on the game performance. (Hypothesis 5; Hypothesis 6).

Method: Participants. 120 participants will be recruited from West Africa (60) and from the USA (60). Each cohort will be split equally between expert (30) and non-expert (30) gamers. An Expert will be anyone who in the last two years had played seven or more hours of video games per week, while non-experts will be those who played one hour or less per week (Boot et al, 2008).

Procedure. A customized VR game will be used for the study. Participants will be sampled at two-time points (game with an ad and without an ad). Ads will vary in duration and frequency. Each gameplay will be one hour long. Visual tests will be administered to eliminate any uncompensated sight challenges. Baseline data that will be collected from participants prior to the game will be demographic information, trait mindfulness, stress level and level of game expertise.

Results: Main and interaction effects will be analyzed via a series of ANOVA and multivariate regression analyses.

Practical Implications: Our study may suggest to marketers and advertisers strategies for effective ad placement within VR games, in a way that will optimize exposure without negatively impacting the gaming experience.

A lower stress level at the end compared to the baseline measure of stress may suggest a contribution from mindfulness, from the gaming activity itself, or both, in the final stress measure.

A higher level of Mindfulness at the end of gameplay compared to the baseline measure may implicate gaming in the outcome. Playing video games can lead to improved levels of some cognitive attributes like attention (Boot et al., 2008), itself an attribute of Mindfulness. This may, therefore, imply that gaming could be a potential alternative to the time-consuming and costly 8-week mindfulness training (Brown & Ryan, 2003), sought by many organizations.

*Eugene Ohu (Lagos Business School, Nigeria)*

### **Applications of Virtual Reality to Mitigate the Effects of Trauma**

Statement of Problem: Children at risk for trauma experience significant impairments across several functional domains in educational settings. Such impairments often involve emotional regulation that impacts interactions (Blair & Raver, 2015). It is currently estimated that 26% of children with in the United States will witness or experience traumatic events before turning four (Van der Kolk, 2017). In high needs schools this number rises to 60% of students (Porche, Costello, & Rosen-Reynoso, 2016). For children experiencing trauma, difficulties within the educational setting manifest as early as Kindergarten and often persist throughout their school career and into adulthood (Hess, Shannon, & Glazier, 2016). The ability to develop strategies to navigate traumatic relapse in classrooms is an area of significant concern. Currently, behavioral modification interventions such as Dialectic Behavioral Therapy (DBT) act as evidenced based approaches for coping with trauma (Miller & Mazza, 2018). However, given the large number of children experiencing trauma in high needs schools, both teachers and students would benefit from low cost, high impact, therapeutic environments making use of the DBT framework through virtual reality (VR).

Procedures: This study builds upon existing research to enhanced teacher-student and student-student focused interactions for children with trauma experiences for use in a school settings. Virtual Reality (VR) enhanced Dialectical Behavioral Therapy treatment program was held for three hours each week over the span of a school year. The program was created based on the premise that VR within the context of DBT would increase the palatability of the program for teachers and increase the engagement for students because of the framing of the treatment, and its soft-failure environments with realistic responses via natural language integration and background analytics collection. In addition, examination of self-perceived trauma reduction responses, physiological response measures e.g. heart rate variability, electroderm activity, and neuroimaging via fNIRS were collected.

Analysis: Statistical analysis was conducted on the standardized hemoglobin absorption ratios between the oxygenated hemoglobin and deoxygenated hemoglobin, examination of participants interview responses, and examination of changes in physiological response both when in the VR scenarios and face-to-face classroom

interactions for 60 at risk students in a high needs urban school. These standardized responses were statistically tested for differences using a Repeated Measures Analysis of Variance (rANOVA) and planned post-hoc comparisons by condition. In rANOVA the subjects serve as their own control making it particularly useful for examining A-B-A within designs in studies such as this one and to identify responses of interest. Specifically those exhibiting responses above Baseline I and II. rANOVA reduces error variance and increases the power of the test to detect differences. The rANOVA was used to assess the main effect of response differences between Baseline I, Stimulus, and Baseline II levels averaged across each condition's participants. Multiple comparisons were done to identify specific differences between Baseline I, Stimulus, and Baseline II levels. One-way ANOVA was used to examine between condition differences on the standardized responses for each condition. In order to reduce the complexity of the data composite data for each optode was used and a Tukey-HSD post-hoc comparison was used to identify statistically different response between conditions.

**Results:** Results of this completed study suggest that the main effect of the treatment condition i.e. VR Based DBT is statistically significantly different in physiological response changes  $F(3,77) = 9.76, p < .001$ . Post-hoc planned comparisons using Tukey-HSD illustrates that the condition of VR reduces trauma responses which are similar to non-VR based DBT over time when directly compared across the variety of physiological measures  $t(38) = 2.98, p < .001$ . Correlational analysis between interview outcomes, physiological responses, and hemodynamics responses suggests a statistically significant relationship between individual simulated events in VR and composite hemodynamic response and physiological response,  $r(78) = .73, p < .001$ .

**Practical Implications:** To be clear, VR technology is not a panacea; the technologies are simply a therapeutic tool. What is different about VR from traditional therapies is the infusion of technology to create environments, and arguably more importantly, to assess therapeutic outcomes. By integrating artificial intelligences in the VR, these systems, the AI allows the VR to become adaptive to a wider spectrum of therapeutic needs. Factors such as memory, skill acquisition, and reasoning ability effect a participant's capacity to incorporate new knowledge into existing prior understandings (Beier & Ackerman, 2005). Incorrect understandings, strategies, and mechanisms need to be confronted in an appropriate manner to help participants develop deeper knowledge and promote automaticity of response. Guided reflection, construction of representations through experimentation, and graphical representations of idealized representations are effective methods in VR.

**Conclusions:** Given the difficulty participants sometimes have in generalizing and applying from one setting to another, the soft-failure environment of VR allows participants to engage with repeated practice of skills and strategies. VR has the potential to be a powerful tool to treat multiple conditions with the flexibility needed by practitioners.

*Amanda Kavner (SUNY Buffalo)*

## Independence D

### Deception and Trust Online

#### PAPER SESSION

#### Teaching College Students the Four Moves of Expert Fact-Checkers

**Introduction:** Recent research indicates that college students fail to evaluate the trustworthiness of online information (McGrew et al., 2018). Unlike expert fact-checkers, students rarely "read laterally." That is, students don't attempt to verify claims made on websites by using fact-checking sites, Wikipedia, or online searches to learn more about a topic and the people and/or organizations making claims about it (Wineburg & McGrew, 2017). We report findings from one of eleven colleges participating in the Digital Polarization Initiative (DPI; American Association of State Colleges and Universities), a national effort to teach college students information-verification strategies commonly used by expert fact-checkers. These strategies rely heavily on using lateral reading for online research (Caulfield, 2017).

**Method:** Participants were first-year college students enrolled in matched sections of a general education civics course at a large urban public university (N = 180, 52.8% female). Students in DPI sections (n = 86) completed online pre-/post-test assessments, received 3 weeks of in-class instruction on fact-checking strategies and online homework assignments to practice skills. Students in control sections (n = 94) followed a standard civics curriculum and completed online pre-/post-tests only.

Pre-/post-tests included four lateral reading problems for which students indicated their level of trust in online photos and articles and explained their reasoning (see Figure 1 for sample problem). Students also completed a scale assessing general and news-related media-literacy knowledge (adapted from Ashley et al., 2013 and Powers et al., 2018) with items such as "Individuals can find news sources that reflect their own political values." They were also asked to report how often they used Wikipedia to verify information and their agreement with statements like "people should trust information on Wikipedia."

**Results and Conclusions:** Students' responses to the lateral reading problems were scored using the following rubric developed by the DPI: 0 (no effort), 1 (reaction or description of content on original webpage), 2 (investigative intent, but did not search laterally), 3 (conducted lateral search but drew incorrect conclusion about trustworthiness of content), or 4 (conducted lateral search and drew appropriate conclusion about trustworthiness of content). At pre-test, both DPI and control groups failed to demonstrate lateral reading (M DPI = 1.23 (SD = 0.45), M control = 1.22 (0.44)). At post-test, students in DPI sections showed gains (M DPI = 2.37 (0.94)), whereas controls did not (M control = 1.13 (0.33)). There was a significant interaction between group (DPI/control) and time (pre-/post-test),  $F(1,178) = 124.18, p < .001, \eta^2 = .41$ .

To examine whether the number of students who made pre/post-test gains differed by group, students' pre/post-test responses were re-coded as 0 if they were at the 0, 1, or 2 level, or as 1, if they were at the 3 or 4 level, and summed. Students whose sums at post-test were less than or equal to their sums at pre-test were coded as 0, meaning that they did not make gains from pre-test to post-test. Students whose sums at post-test were greater than their sums at pre-test were coded as 1, meaning that they did make gains from

pre-test to post-test. Chi-square analyses indicated that significantly more students in the DPI sections (72%,  $n = 62$ ) made gains than would be expected by chance,  $\chi^2(1, N = 180) = 85.70, p < .001$ . Among the DPI students who did not make gains (28%,  $n = 24$ ), 17 did not use lateral reading for any of the problems at pre-test or post-test. By comparison, for the control sections, only 5% of students ( $n = 5$ ) showed gains at post-test. Among the control group who did not make gains (95%,  $n = 89$ ), 77 did not use lateral reading for any of the problems at pre-test or post-test.

Both groups demonstrated high general and news media-literacy knowledge at pre-test (M DPI = 70% correct, M control = 73% correct), with no improvement at post-test (M DPI = 69% correct, M control = 73% correct). This suggests a dissociation between students' media-literacy knowledge and fact-checking skills.

Both groups showed infrequent use of Wikipedia to verify online information at pre-test (M DPI = 2.29 (1.05), M control = 2.22 (0.96)). Students in DPI sections showed gains at post-test (M DPI = 2.58 (1.09)), whereas students in control sections did not (M control = 2.26 (0.97)); interaction between condition (DPI/control) and time (pre-/post-test),  $F(1,178) = 4.30, p = .04, \eta^2 = .02$ . Despite increased Wikipedia use, students in DPI sections continued to report low trust in Wikipedia at post-test.

These findings indicate that the DPI fact-checking curriculum effectively improved students' lateral reading. Increased use of Wikipedia is consistent with the DPI's emphasis on using Wikipedia to research claims and sources. Further research is needed to determine if improvements in lateral reading are maintained over time and explore factors that might distinguish non-responders to the intervention from responders.

*Jessica Brodsky (City University of New York)*

### **Disrupting Democracy: Using Outrage Culture as a Conduit to Spread Propaganda on Social Media.**

An emerging area of study in psychology is the impact of social media on public opinion. Social media has been shown to influence mental health, perceptions of self/others, political beliefs, and a host of other psychological variables (Bright, Kleiser, & Grau, 2015; Fardouly et al., 2015; Bode, 2016; Guntuku et al., 2017). Yet, there has been very little critical analysis concerning the potential misuse of social media by "bad actors."

While propaganda is not a new phenomenon, the scope and subtlety of this influence is new. The full impact of foreign governments, multinational corporations, and activist groups using social media to influence the opinions of citizens in western society has only recently been explored (DeResta, 2018). The laudable qualities of digital platforms—open, free, anonymous—invite exploitation due to precisely these qualities. The very connectivity and intimacy that social media encourages can be used to erode civic discourse. As an example, controversial politically charged stories, videos, pictures, and memes have been created/shared by foreign governments for the purpose of engineering division and increasing "disinformation" in the U.S. (Gerber, 2016; McGeehan, 2018; Stewart, Arif & Starbird, 2018). This type of media is designed to inflame debate, increase misinformation, and even influence elections. The seriousness of this issue has been confirmed by several major social media corporations including Facebook, Twitter and has been substantiated by the US Intelligence Community (United States House of Representatives Permanent

Select Committee on Intelligence, 2017; Facebook, 2018)

The most well documented version of this type of propaganda spreading is that of the Russian Web Brigades, also known as "Russian Troll Farms." The Russian Web Brigades are groups of Russian citizens (generally young tech-savvy individuals) that have been hired by the Russian government to spread misinformation. The goal of such organizations is to "legitimize the Putin regime, garner support for its policies, and demonize its critics. Internationally, they are part of a larger effort to project Russian "soft power," sow doubts and uncertainty within the NATO alliance, weaken public support for policies countering Russian aggression in Ukraine, and solidify the allegiances of Russia's allies in former Soviet republics whom Russia considers part of its natural sphere of influence" (Gerber, 2016 p. 80).

One particularly disturbing characteristic of the Russian Web Brigades is that much of their work has been aimed toward co-opting social justice issues as a means of creating controversy and division (DeResta, 2018) As an example, in 2017 it was revealed that Twitter had identified 2,752 accounts that were linked to Russian Propaganda Efforts (United States, of Representatives Permanent Select Committee on Intelligence, 2017). In one examination, Stewart, Arif, & Starbird (2018) isolated 29 of these accounts and determined that they were utilized to enhance controversy related to race relations, gun violence, and the #BlackLivesMatter Movement. In essence, these accounts were designed to "stoke the flame" of controversy related to the #BlackLivesMatter vs. #BlueLivesMatter/#AllLivesMatter contingencies. This is not only disturbing because of the negative effect of foreign created propaganda on democracy, but also due to the uniquely pernicious method of manipulating reactions to social justice issues as a conduit for spreading this misinformation. Movements that aim to ultimately create a more harmonious society are being used as a tool to increase tension. This appears driven by active manipulation of instances of so-called "outrage culture" online, related to controversial political content. While the Russian Web Brigades are organized and funded by a state, perhaps the most notable feature is how little organization and funding is required to manipulate.

While the technological instruments are complicated, the fundamental psychological dimension is not: this new age of propaganda is driven by well-established psychological phenomena such as confirmation bias and "the bandwagon effect" (Nickerson, 1998; Sundar, Oeldorf-Hirsch, & Xu, 2008). Despite this, very little critical examination of this topic has occurred within the field of psychology. A thorough understanding of the underlying psychological phenomenon is critical in detecting and combating this type of propaganda.

In this presentation, we would like to provide an overview of the research on this type of social media influence, explain how it differs from propaganda of the past, present several examples of social media posts that were produced by Russian Web Brigades, discuss the implications for social justice movements in the United States, and provide recommendations for responsible social media use.

*Kipp Pietrantonio (University of Texas Southwestern Medical Center)*

### **Role of Psychopathic Traits, Cyber Law Awareness, and Cyberspace in Cyberbullying Perpetration and Victimization Among University Students**

A statement of the problem: The empirical research was conducted in legal and psychological context. Cyber-routine activities theory only explains cybervictimization. Hypothetical model was proposed

on the basis of Bandura's moral disengagement and Cyber-routine activities theory to explain the phenomenon of cyberbullying perpetration and victimization by introducing awareness of cyber law and psychopathic traits.

**Procedure:** Formal permission was obtained from authorities of different Pakistani universities of twin cities (i.e., Rawalpindi and Islamabad). The purposive sampling technique was used for data collection. The inclusion criteria of sample was only those university students who were smartphone and internet users. First of all researcher accessed the university students and the purpose of study was briefed. After taking informed consent from students, it was ensured to them that their confidentiality would be maintained and data is used only for research purpose. Due to sensitive nature of statement in the booklet, considerable time was spent in rapport building with students and importance of giving honest responses was highlighted. Thereafter, booklet of questionnaires was distributed consisting of Cyberbullying Perpetration and Cyberbullying Victimization Scales Online Lifestyle Scale, Digital Guardianship Scale, Psychopathy Subscale of Short Dark Triad Scale, Awareness of Cyber Law Scale, and detailed demographic sheet. The administration took place in individual as well as group setting, however, in group setting students were seated away from each other as much as possible to give them a sense of privacy for them. The students were instructed to respond all items and their queries were addressed. After collection of data student and concerned authorities were thanked for doing the needful.

**Analyses:** Data (N =1000) were analyzed through SPSS-22 using descriptive and inferential statistics. The relationships among variables and prediction were established through Pearson Product Moment correlation and multiple stepwise regression analysis, respectively. Structural Equation Modeling through AMOS-22 was used for testing hypothesized model explaining the role of psychopathic traits as predictors and awareness of cyber law, online risky lifestyle, digital capable guardianship, and cyberbullying victimization as mediators of cyberbullying perpetration among university students.

**Results:** Regression analysis revealed that cyberbullying victimization psychopathic traits, male gender and awareness of cyber law are significant predictors of cyberbullying perpetration. Furthermore, awareness of cyber law, online risky lifestyle, and cyberbullying victimization mediated the relationship between psychopathic traits and cyberbullying perpetration; digital capable guardianship was nonsignificant mediator between psychopathic traits and cyberbullying perpetration, and victimization.

**Practical implications:** The present study will be helpful for university students, parents, teachers, and clinical therapist to plan effective strategies for implementing educational opportunities, prevention and intervention programs based on appropriate and secure digital use, awareness, and intervention programs for students in order to prevent or decrease cyberbullying perpetration and victimization. Findings show that will be fruitful for policy makers and computer experts for making computer security software and related applications. Government can devise policies based upon findings of study that how to prevent cyberbullying perpetration and victimization considering awareness of cyber law, secured surfing in cyberspace, and digital capable guardianship. Educationists and IT professionals to design curriculum for awareness of university students about cyber law and secure use of digital media and also provide well equipped computer labs, arrange workshops for awareness, and appropriate use of digital technology.

**Conclusions:** Cyberbullying is a public health problem worldwide. The recent advancement of digital technology in cyberbullying research highlighted the significance of studying awareness of cyber law. Cyberspace have distinct characteristics and most importantly empirical findings suggest that psychopathic traits, risky online lifestyle lead to involvement in cyberbullying perpetration and victimization among university students. The present study was an empirical attempt to fill the literature gap by adding cyberbullying perpetration and other study variables in cyber-routine activities theory.

Cyberbullying is an emerging problem, because information and communication technologies progressively become a part of daily life of university students (Faucher, Jackson, & Cassidy, 2014). More cases of cyberbullying were reported in Punjab, Pakistan (Zarar, 2017). Government of Pakistan has taken steps and passed Prevention of Electronic Crimes Act has been passed by National Assembly (2016) for the prevention of cybercrime and implemented strictly that no one take it as for granted. Therefore, people should be aware of reporting procedure in cyber cell, their role, and responsibilities (Shah, 2016). In Pakistan, previous research suggested and highlighted the unawareness of current cyber legislation against cybercrime, and lack of trust on law enforcement agencies (Magsi, Agha, & Magsi, 2017). Therefore, present research aimed to test hypothesized model focused on Cyber-routine activities theory by Choi (2008) and Bandura's moral disengagement theory (as cited in Naquin, Kurtzberg, & Belkin, 2010). Cyber-routine activities theory only explains the phenomenon of cybervictimization. However, present research fill the literature gap by using two components of cyberspace that is online risky lifestyle and digital capable guardianship in explaining the phenomenon of cyberbullying perpetration and victimization.

*Hira Kanwal (National Institute of Psychology, Pakistan)*

### **Integrating Cognitive Psychology into AI for Fake News Prevention**

While AI solutions have been developed to mitigate fake news on social media, more and more studies revealed that human behavior plays a critical role in the spread of fake news (e.g., [8]). In this work, we seek to strengthen AI approach to prevent misinformation by leveraging insights from cognitive psychology to explore underpinnings of human decision-making. We first describe an empirical user study examining factors that impact individuals' belief in fake news, and link its findings to individuals' decision-making processes. Then, we draw connections between cognitive psychology and AI approaches of misinformation prevention. Specifically, we elaborate on a prototypical example of fake news mitigation using AI, and propose to use insights from human decision-making to inform the AI approach. Finally, we discuss implications for the integration of cognitive psychology and AI approaches.

#### *Cognitive Psychology Approach to Fake News Prevention*

Within experimental settings, a few factors have been investigated to understand how they impact individuals' susceptibility to fake news on social media. Pennycook et al. [5] conducted online studies examining the influence of warning and repetition. Results showed that repeated news headlines were rated as more "real" than novel headlines regardless of headlines' legitimacy and warning. Moreover, the perceived accuracy of news headlines increased linearly with the number of times to which the participants had been exposed.

Pennycook et al. concluded that “the spread of fake news is supported by persistent low-level cognitive processes due to repetition”. Their findings also suggest that individuals might use recognition-based heuristics when they made decisions. A heuristic is defined as a strategy to make decisions more quickly, frugally, and/or accurately than more complex methods by ignoring part of the information [2]. Memory literature reveals that individuals typically have a sense of recognition earlier than that of recollection [6]. So, if individuals recognize one option but not others, they will follow the recognition to make the inference. And such recognition heuristic is typically non-compensatory in that the choice determined by recognition is hard to reverse [3]. If all alternatives are recognized but one is recognized faster, the perceived truth of repeated option will be increased due to the fluency processing from previous exposure [7]. Thus, a fluency heuristic is formulated for situations when the recognition heuristic cannot be applied.

The increased accuracy perception obtained with a single exposure lasted even after a week by Pennycook et al. [5], suggesting that participants may use the recognition heuristic when they made decisions about news’ legitimacy. The perceived accuracy regardless of warning indicates that the decision was hard to reverse, further implying that participants might use a recognition heuristic. Moreover, the perceived accuracy of news headlines increased with the number of exposure, suggesting that participants probably use fluency heuristic.

#### *Integrating Cognitive Psychology in AI: Fake News Mitigation*

Prior work on fake news mitigation deals with competitive influence maximization problems, in which two players are assumed within a social network, one of whom spreads fake news, and the other player mitigates the spread of fake news. Farajtabar et al. [1] proposed a multistage intervention framework, in which they described fake news mitigation as the problem of optimal point process intervention in a social network. Specifically, they modeled the temporal randomness of fake news spreading and mitigation events as Hawkes point processes [4], in which multiple cascades to mitigate against fake news spreaders were allowed.

Unfortunately, there is no prior evidence to suggest that a Hawkes point process is the right model for fake news spreading and mitigation. In this work, we propose utilizing research in cognitive psychology to come up with realistic models for fake news spreading and mitigation. Specifically, we suggest to incorporate human behavior within the nodes of the network. For example, with the role that recognition plays in individuals’ decision-making of fake news (as noted in the previous section), we propose to treat fake news mitigation as the problem of optimal repetition interventions in a social network. Second, we aim to embed these realistic models into AI algorithms to come up with strategies for seed node selection in these game settings, which optimize the extent to which fake news emanating from the fake news spreader can be mitigated. Finally, we aim to evaluate our framework by conducting online user studies to examine whether our proposed AI algorithms can reduce individuals’ susceptibility to fake news, thereby leading to its optimal mitigation.

Conclusion: The spread of fake news on social media is primarily due to decisions made by human [8]. Thus, it is crucial to take human behavior into account while developing AI algorithms for fake news prevention. Our proposed study aim to strengthen AI approach of fake news prevention by integrating insights from human decision-making. Advances in AI algorithms will benefit society by

providing reliable information on social media and increase people’s trust in online information.

*Aiping Xiong (Penn State University)*

4:00–4:30 P.M.

**Independence Foyer**

## Networking Break

---



## KEYNOTE ADDRESS

4:30–5:45 P.M.

**Independence Ballroom**

# The Potential of Policy, Partnerships, and Combinatorial Innovation

In this fireside chat, Kumar Garg, a Senior Advisor for science and technology in the Obama Administration, and now a Senior Director at Schmidt Futures, will discuss why science and technology should be playing a larger role in helping us address our core societal challenges, from education, to health, sustainability, social mobility, and democracy. The conversation will cover a range of issues, from the need for research funding that more deeply integrates the social and behavioral sciences, to harnessing and managing the deluge of data that our digital lives are creating, to the critical role that a range of players—from government, to industry, to research institutions, to the general public—must play in shaping our future.



**KUMAR GARG, JD**

**Senior Director**

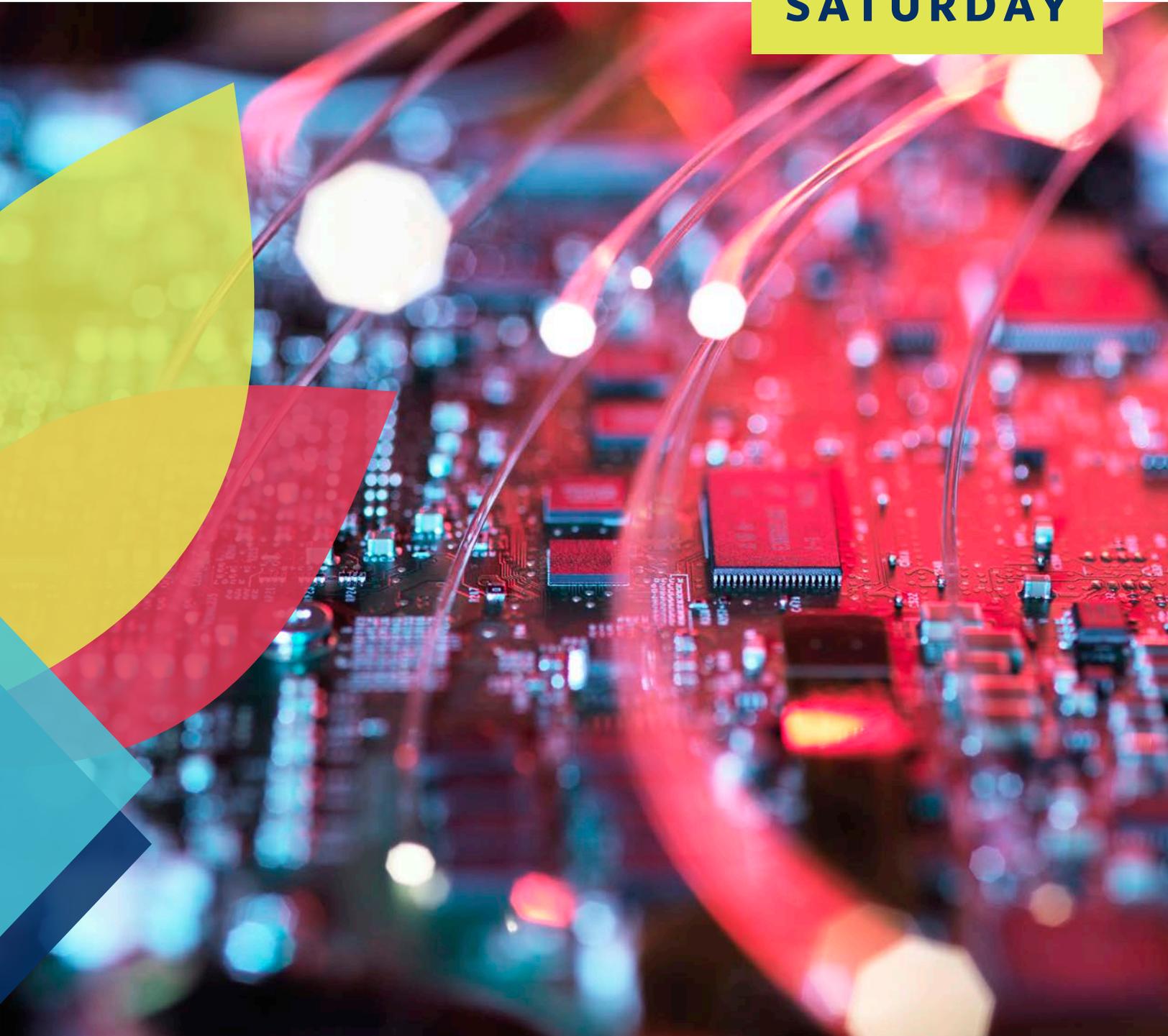
*Technology and Society* Schmidt Futures



**TECHNOLOGY,  
MIND &  
SOCIETY**  
AN APA CONFERENCE

**OCTOBER 5**

**SATURDAY**



7:30-9:00 A.M.

Constitution Ballroom

## Continental Breakfast / Poster Session 2

### Artificial Intelligence

#### A-1

##### The Use of Artificial Intelligence for the Reduction of Symptoms of Depression and Anxiety in Argentine University Students

The use of Artificial Intelligence (AI) in the development of psychological intervention is still in their early stages and further research is needed to proof efficiency of how it can contribute to the mental health field. Tess is a chatbot based on AI that has a high level of scalability since is based in a customize platform. The present study was an RCT comparing chatbot vs. psychoeducation intervention in Argentina. The main goal was to evaluate the effectiveness of the chatbot during eight weeks to reduce symptoms of depression and anxiety. The sample was composed by University students ages 18 to 33 and 87.2% were female, 50 participants completed the depression measures and 63 the anxiety measures. Measures used were The Patient Health Questionnaire (PQH-9, Kroenke, Spitzer, & Williams, 2001) and the Generalized Anxiety Scale (GAD-7 - Newman, et al., 2002). A t-test analyzes for independent samples (experimental group and control group) indicated that there were no statistically significant differences between the study groups in the depression ( $t(48) = 1.47; p = .089$ ) and anxiety scores ( $t(61) = .87; p = .388$ ). However, a t-tests for related samples showed statistically significant differences for the depression ( $t(26) = 2.47; p = .020$ ) and anxiety scores ( $t(26) = 2.15; p = .041$ ) from pre- to post- intervention in the experimental group, and no significant differences were observed for the control group for anxiety ( $t(22) = 1.00; p = .327$ ) and depression ( $t(29) = .413; p = .683$ ).

Preliminary results indicate that using Tess is a promising tool for reducing symptoms of depression and anxiety in University students in Argentina. Further research with larger samples is needed in order to extend understand how chatbots can contribute to psychological interventions.

*Maria Klos (Universidad Adventista del Plata and CONICET)*

#### A-2

##### Engagement and Usability of an Artificial Intelligence Chatbot Supporting a Pediatric Weight Management Program

Background: Pediatric obesity impacts 17% of school-aged children (6-11 years) and 20.5% of adolescents (12-19 years; Ogden, Carroll, Fryar, & Flegal, 2015). Expert guidelines recommend frequent intensive behavioral counseling for youth resistant to counseling from their primary care provider. Patients and families experience barriers to receiving this intensive treatment at the levels necessary to produce significant change, including availability of services, patient retention and dropout due to time conflicts (Hampl et al., 2016; Sallinen Gaffka, Frank, Hampl, Santos, & Rhodes, 2013). Behavioral Technology Interventions (BITs) leverages the benefits of technology to address various health needs through a variety of mediums, such as Internet interventions, video games, mobile apps, and artificial

intelligence (AI). BITs present several possible benefits including increased dissemination and accessibility, cost-effectiveness, increased engagement, and decreased stigma, especially among youth. Tess, an AI chatbot service, addresses various facets of mental health through customized brief conversations via SMS text messaging and Facebook Messenger. Method: This study explored the feasibility and engagement of the chatbot, Tess, which was customized to deliver Behavioral Coaching interventions to support a pediatric weight management program. This pilot featured the novel transfer of in-office tailored goals negotiated with a clinician to the chatbot for further reflection with the patient. Available 24/7, Tess may be able to offer an extension of the office, wherein the patient can access support outside traditional office hours while ideally supporting treatment adherence. Adolescent patients ( $n = 23$ ; Mage = 15.20 years; Range = 9.78-18.54 years; 57% female) coping with weight management and pre-diabetes symptoms enrolled in a weight management program engaged with Tess. Results: A total of 4,123 messages were exchanged between participants and Tess, resulting in about 267 total conversations (with <30 minutes of inactivity). On average, Tess engaged in about 12 conversations per patient ( $SD = \pm 8.84$ ). Longest conversation held was one hour and 13 minutes long, while the shortest conversation fell between four-seven seconds. Average duration of conversations was approximately 12.5 minutes ( $SD = \pm 15.62$  minutes) and the median length was nearly six minutes (00:05:56; IQR = 00:01:54 - 00:17:32). Overall, Tess provided about 55 hours and 45 minutes of support for the adolescent patients (over a specific span of time), 17.8% of which was provided outside of typical office hours (8am-5pm). Conclusions: The AI chatbot, Tess, was successfully implemented as an additional support in a weight management program at a children's hospital. The adolescent patients engaged with the chatbot and reported their interactions as useful. These results highlight the feasibility and beneficial support that this AI chatbot can offer, specifically in a pediatric setting, which could be scaled to serve larger groups of patients. Tess was able to continue the therapeutic interactions outside office hours via a preferred method of communication, all while maintaining patient satisfaction. Technology may be uniquely situated to enable patients to obtain more personalized and on-demand care.

*Taylor Stephens (Palo Alto University)*

#### A-3

##### Applicant Reactions to the Use of Artificial Intelligence in Selection

Recruitment and selection are the initial stages of the staffing process in which organizations seek to attract a large number of qualified applicants. In addition to motivating targeted individuals to apply, this also includes whether applicants will stay in the pool, and whether they will accept an offer. With these goals in mind, organizations must consider whether the targeted individuals have favorable

reactions to the way they are treated by the organization during this process.

Artificial intelligence (AI) is increasingly being utilized by organizations in making selection decisions. For example, technologies such as resume screening and interview assessments by artificial intelligence (AI) allow companies to process large numbers of applications in a short time, and help them save time and money. Because of these benefits, more organizations are embracing AI as a tool in staffing (Alsever, 2017; Florentine, 2017). However, the rate of adoption of these technologies has outpaced research, and it has been recommended that these technologies be utilized with caution, especially in regard to employment laws (Barocas & Selbst, 2016). In addition, research has yet to examine whether applicants have favorable reactions to an increased involvement of AI in the selection process, which has been shown to influence a number of important organizational outcomes (Gilliland, 1993; Hausknecht, Day, & Thomas, 2004).

Applicant reactions refer to attitudes, affect, or cognitions an applicant might have about a hiring process (Ryan & Ployhart, 2000). Generally, applicant reactions consist of procedural and distributive justice perceptions, attitudes and motivations towards test[s], and test anxiety (Hausknecht et al., 2004; McCarthy, Bauer, Truxillo, Anderson, Costa, and Ahmed, 2017). Distributive justice is the perception of fairness of the outcomes of organizational decisions, whereas procedural justice is an aspect of the process through which decisions are made, focusing on the quality of interpersonal treatment people receive during the duration of organizational procedures. In the selection context, perceptions of procedural justice would include (a) job relatedness of procedures, (b) opportunity to perform, (c) reconsideration opportunities, (d) consistency, (e) feedback, (f) provision of justification for a decision, (g) honesty, (h) interpersonal effectiveness of administrators, (i) two-way communication, and (j) propriety of questions asked during selection (Gilliland, 1993). Accordingly, two processes may shape applicants' reactions to the use of AI in selection; (a) applicants may view AI favorably as they believe it will reduce human subjectivity, or (b) applicants may believe that the lower amount of human interaction means the organization does not value them.

The purpose of this study is to investigate the effects of AI utilization in selection on applicant reactions. Depending on which one of the above processes is more dominant in shaping reactions to the use of AI, we expected that applicants would either have positive or negative reactions, which would be observed through perceptions of procedural and distributive justice, overall organizational attraction, job pursuit and recommendation intentions, and litigation intentions. In addition, we hypothesized that outcome favorability would moderate the relationship between AI use and justice perceptions such that the relationship would be more positive for accepted applicants. Finally, it was hypothesized that self-efficacy with computers would moderate the relationship between the use of AI and perceptions of justice such that the relationship would be more positive for those with high levels of self-efficacy.

The data for this study was collected through an online survey (N = 320) in which participants were randomly presented with one of four vignettes, describing a hiring scenario in which the decision-maker (human vs. AI) and the outcome (accept vs. reject) were manipulated, leading to a 2x2 experimental design. Participants then answered survey items centered around (a) justice perceptions, (b) organizational attraction, (c) job pursuit intentions, (d)

recommendation intentions, (e) litigation intentions, and (f) self-efficacy with computers. Findings indicate that reactions to the use of AI in selection decisions are generally negative with AI use being negatively correlated with justice perceptions and organizational attraction outcomes and positively correlated with litigation intentions. The results also support the hypothesis that outcome favorability moderates this relationship with the effect being significantly stronger for accepted applicants than rejected applicants. No support was found for the hypothesis that self-efficacy with computers would moderate the effect. Findings from this study will have important implications, including guiding organizations on how best to utilize AI in selection in order to attract and retain top talent through the recruitment process.

*Maira Compagnone (Appalachian State University)*

#### A-4

##### **"Jub Jai" : A Thai AI Chatbot for Depression Screening on Facebook Messenger**

Depression is among the most concerning mental health problem globally including Thailand. Depression rate in Thai population has been increasing rapidly but the awareness level in this emotional disturbance is not high (Kongsuk, 2016). This is due to a limit number of mental health professions, fear of being stigmatized from seeing mental health professions as well as lack of understanding in depression (Thai Depression Centre., 2016). In order to alleviate this problem, it is important to make an easy to access early depression detection available to encourage Thai people to assess their depression level regularly.

Facebook is reported as the most popular social media platform in Thailand. Currently, there are over 2.23 million active Thai users (Pornsakulvanich, 2018). Because of this, authors sought to apply psychological knowledge with artificial intelligence to develop a friendly Facebook Chatbot for depression screening. This innovative research project aimed to provide an AI chatbot that makes users feel free and more open to talking with than answering a traditional self-rating depression questionnaire.

To develop a friendly depression detection chatbot, several psychological research procedures were followed. Firstly, 20 depression items from the Thai Mental Health Questionnaire (TMHQ), a Thai standardized mental health inventory, was factor analyzed (Phattharayuttawat, Ngamthipwattana, & Sukhatungkha, n.d.). Exploratory factor analysis was employed. Data from over 600 participants provided a 9-item version of TMHQ depression scale with 4 factors; anhedonic, sleep problem, cognitive function and suicidal ideation. A total score was used to calculate mean and standard deviation in order to classify depression level into 4 levels: normal, low, medium and high.

Next, the 4-factors of depression scale with 9 items were then transformed to a conversation dialogue structure. To clarify, the 9 items remain unchanged. Users still had to rate how much each item describe them but other conversations were added. For instance, instead of asking about whether the user had a problem with sleeping, JubJai chatbot would chat about sleeping in general first then ask the user to rate to what extent they had a problem with sleeping. Another example is, instead of asking if the user still enjoy doing their hobbies or daily life activities, the conversation about the user's plan over the weekend or their favourite leisure activities were added before asking

the user to rate their level of enjoyment in doing daily life activity. To ensure that the conversation dialogue is unbiased or misleading, this process was supervised by a licensed clinical psychologist.

Then AI was implemented. This implementation of AI was to make the system understand the user's emotional expression. Natural Language Processing (NLP) is a crucial part of this process. Authors used an algorithm to classify a sentence into three emotional categories: positive, negative, and neutral. This is a novel algorithm since, to the authors' best knowledge; there has not been an emotional classification algorithm in the Thai language. After that, a suitable conversation flow was selected based on the result of the NLP algorithm. Depression score and level were calculated from the main 9 depression items.

In order to examine the psychometric property of this chatbot in comparison to the original 20-item depression scale, the authors called for participation from Facebook users. Over 20,000 people took part in this research. Participants were asked to chat with Jubjai on Facebook messenger and answer TMHQ-depression scale. Results reveal high correlation between total score from TMHQ and Jubjai chatbot (Cronbach's  $\alpha = .88$ ,  $p < .01$ ). Additionally, users reported more satisfaction chatting with AI chatbot than answering the TMHQ-depression scale. The results indicated that Jubjai facebook chatbot can be used as an easy to access and friendly tool for depression screening. This can also be seen from the number of Jubjai users that has reached over 100,000 users in one year.

*Panida Yomaboot (Mahidol University, Pordeekum; AI Company Limited)*

## A-5

### Gender and Racial Bias in Artificial Intelligence systems: Approaching Inclusiveness and Diversity in AI

Artificial intelligence systems are increasingly being adapted to human activities. As a result of the endless interplay between society and technology, human beings are mapping their human biases into the systems build, algorithms, data use to train those systems. Moreover, much of the data currently available is dominated by white, educated and rich men. This is mainly due to their high status in the social hierarchies of the past and present, as well as their high investment in and decision-making power over digital technologies.

Machine learning systems trained on the basis of these data reproduce this bias in their decision-making behaviour. However, this abstract reflection of socio-political decision-making behaviour in intelligent systems can offer society the opportunity to identify and deconstruct their occurrence, to question their meaning and to change them. Currently, academia, industry, governmental forces of different countries are focusing their attention on discrimination and exclusion caused by algorithms and AI systems, directing much of their interest in governing AI, approaching it from societal, legal, philosophical, psychological, ethical perspectives. Tech giants focusing on de-biasing algorithms, fair algorithmic decision-making, bringing the explainable component to black-box algorithms and building trust between human and AI systems. The public sector also directing its vision towards creating an inclusive digital society by involving citizens into the discourse analysis, creating interdisciplinary and inter-cultural research environments. The AI strategies of developed countries include the way AI is used for a common good and reducing inequalities in society, as well as supporting low-resources

communities. In parallel to that, there is a value on researching from where inequalities in AI systems come from, under which framework it impacts the society, and which consequences it has on a long-term basis. With respect to that, the questions need to be addressed are:

How can the innovative added value of machine decision behaviour be used and legitimised in as many areas as possible in order to expand the scope of human capture and action? How can the hierarchy between the system decision maker and system developer be deconstructed by unequally distributed knowledge about digital technologies? How can a social consensus be reached on new standards for intelligent systems in order to place their development as a democracy-friendly process under the responsibility of all individuals?

In the developed concept space which is taking into consideration, users, developer/provider, machine/algorithms, impacted people and user perspective we are discussing the interplay between AI and society and how the gender and racial biases in AI should be approached and governed.

*Gunay Kazimzade (Technical University of Berlin)*

## Childhood / Developmental

### B-1

#### Female Intelligent Media Characters Increase Early Math Skills for Gender-Typed Girls

Statement of Problem: Science, technology, education, and mathematics (STEM) are stereotyped as male domains.<sup>1</sup> Media depictions of STEM often reinforce this stereotype and transmit cultural expectations about gender to children.<sup>2</sup> Girls' endorsement of these stereotypes has been associated with decreased interest and performance in STEM tasks.<sup>3</sup> Providing children experiences with counter-stereotyped competent females in STEM may help to engage girls in STEM and diversify both boys' and girls' gender-STEM associations. Girls may be more exposed to counter-stereotyped messages compared to boys, since boys are more constrained by cultural ideas of masculinity.<sup>4</sup>

Preschool-aged children often use gender to guide their cognitions and behaviors and employ gender stereotypes, a simple type of gender schema, during new experiences.<sup>5</sup> As children learn how society defines gender, this can be organized using these gender schemas to: 1) evaluate whether the information should be placed in "male" or "female" categories and 2) determine how this information will be integrated, or not, into thoughts and behaviors for the self.<sup>6</sup>

Intelligent agents are a recent development in children's media<sup>7</sup> that can be designed as stereotyped or counter-stereotyped embodied characters. Counter-stereotyped intelligent agents may be a useful tool to engage children in diverse STEM experiences. For example, a socially meaningful female Intelligent Character aided both boys and girls in mastery of the add-1 rule,<sup>8</sup> knowledge that adding 1 to a number increases the total by a single unit.<sup>9</sup> The purpose here was to explore whether preschoolers' gender-stereotyped preferences influenced learning of the add-1 rule in interactions with a same-or opposite-sex Intelligent Character.

Procedure: Within sex groups, 82 children (Mage = 4.48 years, SD = .32; 42 girls) were randomly assigned to a screen-based interactive game featuring Dora from Dora the Explorer or Diego from Go

Diego Go! Children gathered birthday party supplies with the character by solving three rounds of add-1 problems (1+1 through 4+1) as supplies came down a conveyor belt. Items increased in difficulty across rounds. The Intelligent Character provided contingent feedback to the child, including scaffolds for incorrect responses, via an experimenter operating a computer behind a room divider. Children's latency scores in answering math problems correctly were measured in seconds from videos of the experimental game play sessions.

After game play, children's gender-stereotyped toy preferences were assessed with a behavioral task where children sorted feminine (e.g., doll, tea set) and masculine (e.g., truck, tool set) gender-stereotyped toys based on how much they wanted to play with each toy.<sup>10</sup> Children's responses were scored, 2 ("a lot"), 1 ("a little"), or 0 ("not at all") points, and average composite toy preference scores were calculated for the feminine and masculine toys separately.

Analyses & Results: Table 1 presents frequencies of toy preferences for feminine and masculine toys. Girls were more likely to report a preference for feminine toys compared to boys,  $X^2(2, N = 82) = 30.19, p < .001$ . Boys were more likely to report a preference for masculine toys compared to girls,  $X^2(2, N = 82) = 16.20, p < .001$ .

An OLS regression was conducted predicting average latency scores by child-character sex match, feminine toy preference, and their interaction in the Dora condition (Table 2). Girls with a stronger preference for feminine toys correctly answered math problems faster than boys ( $p = .03$ ). Boys with stronger preferences for feminine toys were slower at answering math problems correctly ( $p = .001$ ). The same OLS regression in the Diego condition yielded no significant predictors.

OLS regressions predicting average latency scores by child-character sex match, masculine toy preference, and their interaction in both the Dora and Diego conditions yielded no significant results.

Practical Implications and Conclusions: This study contributes novel information about how children's gender-typed preferences influence learning STEM learning from an Intelligent Agent. Stronger preferences for feminine toys facilitated mastery for girls but slowed performance for boys when a female Intelligent Character was featured in the game. The findings suggest that girls' gender schemas may be altered when interacting with a same-sex Intelligent Character in a STEM domain if they have strong feminine preferences. Girls may view interactions with a female character as salient for themselves and be motivated to participate and, in turn, demonstrate higher mastery of the add-1 rule.

However, for boys with stronger feminine interests, interacting with a female Intelligent Character in a typically male-stereotyped field may activate gender schemas in competing ways. When boys with feminine interest encounter the female Intelligent Character engaging with them about math, it may take these boys additional time and cognitive resources to determine how this experience will be categorized in their gender schema.

The results have implications for the design of educational technology products and suggest that girls can benefit from interacting in a stereotypically male domain with a female Intelligent Character. For boys with non-traditional interests, however, a different approach may be needed.

*Marisa Putnam (Georgetown University)*

## B-2

### Same-Sex Intelligent Media Characters Increase Young Children's Early Math Skills

Statement of Problem: Early math skills are foundational for, and predictive of, children's academic achievement. The add-1 rule, knowing that adding one to a number increases the total sum by a single unit, facilitates children's mastery of more complex math tasks. While both girls and boys demonstrate similar basic numerical competencies, science, technology, engineering, and math (STEM) is often stereotyped as a male domain.

Increasing interest and skills in STEM for both boys and girls in childhood may help to engage children in STEM domains. Gender schema theory, which involves organized networks of cognitive associations between gender and self that help children process information, may facilitate children's engagement with STEM content if they consider it relevant to their gender. The traditional dichotomy of gender is particularly salient among preschool-aged children as an organizer of daily experiences.

Intelligent Characters, a type of Intelligent Agent, may serve as effective 21st century mathematics teachers when embodied as socially meaningful media characters. Children have socially contingent parasocial interactions (PSIs) with characters they see onscreen, where the character asks a question, pauses for a reply, and responds as if they heard what the child said. Socially contingent math PSIs with a socially meaningful female Intelligent Character improved boys' and girls' accuracy in answering add-1 math problems faster, indicating increased skill mastery. The purpose here was to explore whether preschoolers' learning of the add-1 rule is influenced by interactions with a same-or opposite-sex Intelligent Character.

Procedure: Within sex groups, 82 children (Mage = 4.48 years, SD = .32; 42 girls) were randomly assigned to a screen-based interactive game featuring Dora from Dora the Explorer or Diego from Go Diego Go! Children gathered birthday party supplies with the character by solving three rounds of add-1 problems (1+1 through 4+1) as supplies came down a conveyor belt. Items increased in difficulty across rounds. The Intelligent Character provided contingent feedback to the child, including scaffolds for incorrect responses, via an experimenter operating a computer behind a room divider. Children's math parasocial interactions were measured by dividing the number of on-task math responses (i.e., numerical replies) by the number of prompts available. Children's latency scores in answering math problems correctly were measured in seconds from videos of the experimental game play sessions using Noldus the Observer XT.

After game play, children completed a transfer task about the add-1 rule by completing 10 addition problems (1+1 - 10+1) with physical objects. In the task, children were presented with an initial number of stickers that were placed in a bag, then presented with one more sticker, and asked what the total number of stickers was. The number of transfer task problems children answered correctly was recorded.

Analyses & Results: Table 1 presents means and standard deviations for all results. An OLS regression was conducted predicting how long it took children to answer the twelve in-game math problems correctly by child-character sex match, math PSI, and their interaction (Tables 1 and 2). Although children that engaged in more math PSI with either character correctly answered more math problems faster, children who interacted more with a same-sex character were significantly faster than children who interacted more with an opposite-sex character ( $p = .03$ ).

An OLS regression predicting math PSI by condition and child sex revealed that children had more math PSI with Diego than with Dora ( $p = .02$ ), when controlling for the sex of the child (Tables 1 and 3).

An OLS regression was conducted on the number of transfer problems children answered correctly predicted by child-character sex match, math PSI, and their interaction. Children that interacted more in the game with a same-sex character, compared to an opposite-sex character, performed significantly better on the transfer task ( $p = .001$ ; Table 1 and 4).

**Practical Implications and Conclusions:** This study contributes to our knowledge of how interactions with Intelligent Characters influence young children's early learning of math skills. Socially contingent parasocial interactions facilitated add-1 performance in the onscreen task for all children, but resulted in the quickest correct answers when children engaged with a same-sex character. The findings suggest that increased interactions with a same-sex Intelligent Character may activate gender schemas which guide learning and memory. Children interacting with a same-sex character may view these interactions as more salient for themselves, which may explain why these children demonstrate more mastery of math problems in the transfer task as well. However, children's engagement in more math PSI with a male character overall suggests that children may view a male talking about math as more credible than a female character, in line with stereotypes about STEM domains. The results have implications for the design of educational technology products and suggest that the use of interactive same-sex Intelligent Characters may benefit STEM learning as early as the preschool years.

*Marisa Putnam (Georgetown University)*

### B-3

#### **Testing the Levels of App-Based Cognitive Rehabilitation Program for Youths with Neurodevelopmental Disorders: A Pilot Study**

Persons with neurodevelopmental disorders (ND) show significant deficits in various areas of executive functioning. Applied Behavior Analysis (ABA) is an evidence-based treatment for many of these disorders but has several limitations to be practiced in real-life setting. It is an expensive service which requires intensive involvement of professionals and paraprofessionals in terms of time, space and other resources. Moreover, lack of specialized institutions and professionals hinders the treatment to be reached out to children and their families who desperately need the intervention. Considering these limitations and recent trends of integration of ICT technology into mental health treatment, new technology-based interventions are emerging as alternatives to traditional face-to-face treatment. Diverse treatment strategies developed based on ABA principles, such as reinforcement schedule, repeated instructions, antecedent-based strategies, etc, could be easily be applicable in technology-based interventions. Adding to this, high interests for IT devices among youths with neurodevelopmental disorders could be contributed to successful implementation of technology-based interventions.

The Y On!, an app-based cognitive rehabilitation program for children with neurodevelopmental disorders, was developed to facilitate four sub-areas of executive functioning; flexibility, working memory, inhibition, and planning. The Y On! is consisted of 24 games, 6 games per each area, and each game has 15 difficulty levels. The Y On! includes various elements of ABA, which are effectively implemented

with technology in the application. Multilayered reinforcement system along with diverse design and structure components, which were identified as successful game factors in the gamification literature, is provided automatically at appropriate timing according to children's performance. Appropriate prompts also appear repeatedly throughout the game via 3-step instruction method, assisting the children when they have difficulties performing the game. Assessment stage at the start of the program determines each user's level of executive functioning and thus offers the appropriate level of the game to begin with according to algorithms built in the application. This individualized offer of the service enriches the efficiency of the program in the sense that users can start with necessary level of training right away. Finally, the program is all data-based. Since the whole program is managed by technology, correction rate and reaction time of the users in each trial can easily be scored as data. Based on the performance level calculated upon the data, appropriate feedbacks are provided to each user.

The purpose of this study was to test the adequacy of difficulty level in a few games on the Y On!. Among the 24 games, 6 games; three from working memory, and one from each flexibility, inhibition and planning were pilot-tested. 12 qualified participants with neurodevelopmental disorders (intellectual disability(8), ADHD(1), ASD(3)), aged 8-13 years performed the skeleton tasks. The correction rate and reaction time of each participant were measured as variables to rate the performance level of the users and evaluate the difficulty of each level. To enhance the efficiency of testing, three levels with a significant change in difficulty were selected and tested in each game.

For analysis of the data, Kruskal-wallis test was carried out to compare the measured correction rate and reaction time between the levels. Kruskal-wallis test, a nonparametric test, was adopted to compensate for the dissatisfaction of assumption of normality, since the number of samples was too small. If the data of the samples within a game met the assumption of normality, one-way ANOVA was carried out subsequently to earn more reliable results.

As a result, there were significant changes in the correction rate as levels got higher in 5 out of 6 games: Play the music, Train, Matryoshka, WCST, and Supermarket. Especially, WCST had the most reliable result showing significant change across the levels through parametric test. However, changes in reaction time showed no significant results in any of the games. Findings suggest that difficulty of each level throughout the games are valid but needs more modification in order to increase sensitivity. Limitations and further implications are discussed.

*Seungmin Jung (Yonsei University)*

### B-4

#### **Media Healthcare in Action: The Center on Media and Child Health Clinician Toolkit**

Today's adolescents live in a world dominated by smartphones and interactive media. This 21st century environment comes with everyday challenges such as negotiating rules about technology use with parents and balancing entertainment and social activities with other responsibilities. For some teens, this environment becomes an unmanageable imbalance of activities characterized by high levels of interactive media use, low levels of non-media activities, and subsequent harm to their social, familial, physical, and mental wellness.

Such disordered media use often occurs with comorbid mental health issues and can trigger acute anxiety, depression, and homicidal and suicidal behaviors.

Currently, primary care physicians (PCPs) are grossly underprepared to identify adolescents at risk for experiencing Problematic Interactive Media Use (PIMU); have limited access to brief, comprehensive, up-to-date, validated screeners for dysfunctional gaming, binge-watching, and problematic social media use or pornography; and need strategies for providing care or proper referral to mental health professionals.

Recognizing the need for a rubric able to effectively help clinicians identify and treat patients suffering from PIMU, in 2014, the Center on Media and Child Health at Boston Children's Hospital launched the Clinician Toolkit to help clinician's optimize the level of care they provide their patients and their families.

Continually evolving based on the latest science from the field and a landmark pilot study conducted in 2018, the Toolkit is designed to help clinicians talk to their patients and their patients' parents about their media exposure and use. Additionally, the Toolkit contains the research, guidance, and materials clinicians need to inform patients about the effects of media on their health, instruct them on how to best use media, and set them up for success by providing guidance, tips and tools that they can use in their everyday lives.

The focus of this presentation will be an examination of the newly redesigned Toolkit following a 2018 pilot study, its reception during a 2019 feasibility study within a clinic, and an examination of the study results based on survey data examining implementation, ease-of-use, and perceived effectiveness.

Procedure: The goal of the CMCH Clinician Toolkit is to provide clinicians with the understanding, guidance, information and streamlined tools they need to care for today's pediatric patients and their families. In order to ensure Toolkit's optimal use and effectiveness, this project will focus on a comprehensive testing program carried out in three phases:

Phase 1: Selection and Training. All Harvard Pilgrim Healthcare member pediatric practices were eligible to apply to participate in the Clinician Toolkit study. Applications were reviewed and a practice was selected based on their eligibility and ability to participate.

Clinical and administrative employees of the selected practice were provided with an in-house comprehensive training and workshop seminar led by CMCH experts and Dr. Michael Rich, Founder and Director of CMCH, Pediatrician at Boston Children's Hospital, and Associate Professor at Harvard Medical School and Harvard T.H. Chan School of Public Health. The CMCH team explained, trained and led all workshop attendees through the Clinician Toolkit, and provided background information about media-related health concerns. Clinical and administrative staff were also given access to all materials needed to effectively integrate the Toolkit into their everyday pediatric practice.

Phase 2: Testing. During this phase, the selected practice will implement and use the CMCH Clinician Toolkit with pediatric patients and their families. Reporting on Toolkit's use and effectiveness will be carried via brief online surveys on an ongoing basis throughout the 2 week testing period. During this time, Dr. Rich and the CMCH team will be available for support and problem-solving for all aspects of testing.

The CMCH team will also welcome any additional anecdotal feedback from any member of the participating HPHC practices. Feedback will be recorded and additional support offered when needed.

Phase 3: Data Analysis and Toolkit Revision. Following completion of the testing phase in April, 2019, the CMCH research team will perform a statistical analysis on all data and feedback collected throughout the study. This data will be collected, translated and organized into a comprehensive report in order to make further improvements to the Clinician Toolkit.

Using the reported data and feedback the CMCH team will revise the Toolkit in order to improve both its ease of use and its effectiveness. This data will also be used to inform further study of the Toolkit's effects on patient health and well-being.

*Kristelle Lavallee (Boston Children's Hospital)*

## Cyberlearning

### C-1

#### **Understanding creativity, attention, flow, feedback, and learning using an integrated electroencephalography (EEG) and immersive virtual reality (VR) system**

In a series of three experimental research studies, we used an integrated system of electroencephalography (EEG) and immersive virtual reality (VR) to study the relationships between creativity, attention, flow, feedback, and learning in an immersive virtual reality setting.

1) In the first study, we explored the connections between an individual's creative behavior and his or her creative brain by asking a total of 65 U.S. museum adult visitors each to design an open-ended virtual product in an integrated system consisting of virtual reality and brainwaves.

The results showed a significant correlation between the individual creativity level and the state of flow. There is a significant correlation between the state of flow and the quality of the creative product. The attention value is correlated significantly with the quality of product. The sequential analysis showed that people with high product creativity had a higher attention value while maintaining a state of relaxation.

2) In the second study, we explored more in-depth the effects of immersive VR on an individual's creativity and factors related to creativity including flow, attention and meditation (i.e., mental relaxation or stress) of 60 undergraduate students in China. The students were invited to participate in an open-ended challenge to design a wearable technology functioning as a smart phone individually, either in the immersive VR condition (N = 30) or in the paper-and-pencil condition (N = 30). Surveys and electroencephalography (EEG) measurements were used to collect data. Findings showed that the participants in the immersive VR condition had higher quality creative products than those in the paper-and-pencil condition. The sequence analysis of the participants' EEG brainwaves also showed that the students in the immersive VR condition maintained a more stable focus or attention, while those in the paper-and-pencil condition were more relaxed. Research results led to a deeper understanding of using a VR support system for individual creativity.

3) The third study investigated whether the feedback designed based on EEG signals contributed to an individual's creative performance in an immersive VR setting with 60 high school students. Two specific forms of feedback were used. The first one was "reminder feedback," given when brainwaves indicated the participant's attention was not concentrated. The second one was "encouraging

feedback,” given when brainwaves indicated that the participant’s attention was very concentrated. Twenty (N=20) participants received no feedback; twenty (N=20) participants received reminder feedback; and another twenty (N=20) participants received encouraging feedback. Findings showed that the participants who received reminder feedback had higher-quality creative products than those in a group with no feedback or encouraging feedback. In addition, EEG feedback also had an impact on the participant’s attention and flow state. These findings are discussed in terms of feedback impacts, study limitations, and future research directions.

*Lin Lin (University of North Texas)*

## C-2

### The Impact of Gesture and Distributing Story between Multiple Characters on Children’s Engagement with an Animated Story Telling Application

Effective storytelling relies on engagement and interaction. This work develops an automated software platform for telling stories to children using animated characters. It experimentally investigates the impact of two design choices on children’s engagement and willingness to interact with the system: story distribution and the use of complex gesture. A storyteller condition compares stories told in a third person, narrator voice with those distributed between a narrator and first-person story characters. Basic gestures are used in all our storytellings, but in a second factor, some conditions are augmented with gestures that indicate conversational turn changes, references to other characters and prompt children to ask questions. The system was evaluated using a 2x2, within-subjects experimental design in which every child saw every condition. A total of 33 students, 12 girls and 21 boys, ages 5 to 8, partook in the study. The stories were all based on Aesop’s fables and contained two or three characters, including the narrator. Output measures include annotated gaze behavior and an analysis of questions children asked the system at the end of the story. An analysis of eye gaze indicates that children attend more to the story when a distributed storytelling model is used. Gesture prompts appear to encourage children to ask questions, something that children did, but at a relatively low rate. Approximately half the children asked one question and 27 total questions were asked for 132 story tellings. Interestingly, the children most frequently asked “why” questions. Gaze switching happened more quickly when the story characters began to speak than for narrator turns. These results have implications for future agent-based storytelling system research.

*Michael Neff (University of California, Davis)*

## C-3

### Computational Thinking with Zoombinis: Results from a national classroom implementation study

Students in classes where teachers used bridging (i.e., making connections between the implicit learning in a game and formal learning in the classroom) showed greater improvement on pre-post assessments than students in classes that had no bridging (Rowe et al., 2014; Asbell-Clarke, Rowe, Bardar, & Edwards, under review). More recent research applies this bridging model with the construct of computational thinking (CT). In a national implementation study of the logic puzzle game Zoombinis, 681 students in 54 elementary and

middle school classes and teachers were provided with classroom materials designed to help bridge implicit game-based knowledge with formal instruction. This paper reports on the multilevel analysis of the role of game-based pedagogy on changes in students’ pre-post assessments.

Zoombinis (TERC, 2015; Hancock & Osterweil, 1996) is a series of challenging logic puzzles (Figures 1a-1c). The game mechanics in Zoombinis align remarkably well with today’s foundational CT constructs of Problem Decomposition, Pattern Recognition, Abstraction, and Algorithm Design (CSTA, 2017; Shute, Sun, & Asbell-Clarke, 2017).

This study explores the question: How are changes in students’ CT assessments impacted by variation in Zoombinis game-centered pedagogy, after accounting for teacher, classroom, and student-level demographics? During the 2017-18 academic year, 36 US teachers teaching 54 3rd-8th grade classes participated in this study. Students in each class completed an 18-item pre-post CT assessment using digital interactive logic puzzles aligned with foundational CT constructs—Problem Decomposition, Pattern Recognition, Abstraction, and Algorithm (Asbell-Clarke, et al., under review). Three types of metrics were used as assessments from these tasks—percentage correct; number of moves; and number of runs (attempts). A total of 681 students (429 elementary, 252 middle-school) completed both assessments and are included in these analyses.

During the study, teachers were asked to log their CT instructional practices and their classroom use of Zoombinis. We are currently creating four major categories of features based on these 938 teacher logs and gameplay data (Almeda et al., 2019). Sample features include: 1. Computational Thinking Instruction; 2. Game-centered pedagogy; 3. Zoombinis gameplay.

In addition to teacher logs, we collected the following types of data from teacher application surveys: teachers’ prior Zoombinis usage, teacher demographics (e.g., number of years of teaching experience), classroom characteristics (e.g., course level—advanced/Honors vs. not; subject area), and school demographics (e.g., percent of students receiving free or reduced-price lunch. Teachers also provided student demographics (grade level and gender).

Using the SPSS MIXED linear models procedure, we first estimated an unconditional, hierarchical 3-level model with students, classrooms, and teachers using Maximum Likelihood (ML) and unstructured covariances for the four CT constructs. This poster will present the best-fitting model of for each CT construct.

This study represents a first step towards replicating our bridging model to a broader context of games and CT instruction, providing insight on the impact of bridging on changes in students’ CT practices. Consistent with our previous work on bridging in two physics games (Rowe et al., 2014; 2017), features from this study will be used for multilevel analyses to examine the relationship of Zoombinis bridging activities on changes in students’ pre- and post-assessments scores on computational thinking. Overall, our research has potential implications for incorporating games in teacher professional development and supporting teacher bridging to improve explicit STEM learning.

*Elizabeth Rowe (EdGE at TERC)*

**C-4****Exploring the Engineering Design Process to Engage Students in STEM and Cyber**

The nation will require approximately one million more science, technology, engineering, and math professionals than what will be produced at current rates over the next decade. The National Integrated Cyber Education Research Center (NICERC), the academic division of the Cyber Innovation Center, is focused on growing and educating the next generation cyber-literate workforce through K-12 education. NICERC curricula fosters creative problem-solving skills while providing real-world cyber scenarios as context. By using the Engineering Design Process in classroom curricula, NICERC content keeps students focused on the steps required to solve challenging, real-world problems using STEM, cyber, and even computational thinking and computer science skills.

*Chuck Gardner (Cyber Innovation Center)*

**Health Care / Telehealth / Digital Health****D-1****Reliability of Drone Observations for Recording Child Behavior on School Playgrounds**

Research in school psychology has made recent advancements in the efficacy of video conferencing software (VC) in observations of students during evidence-based interventions. In-vivo systematic direct observation (SDO) has only recently been compared to SDO using videoconferencing (VC) software, showing that VC SDO is just as effective in estimating student on-task behavior as in-vivo SDO (Fischer, Dart, Lehman, Polakoff, Wright, 2018). Techniques within SDO include momentary time sampling (MTS), partial interval, and whole interval in order to assess student behavior (Cooper, Heron, & Heward, 2006). Because SDO can be difficult to implement (Brown-Chidsey, 2005), particularly within group settings in a classroom, integrating technology in these observations, like VC should be proposed to maximize efficacy. In the context of school-based interventions, technology telehealth training has been found to be as effective as traditional face-to-face applied behavioral analysis training in teachers (Tomlison, Gore, McGill, 2018). Now, further analysis of various other applications of observation technology is necessary.

One area technology enhanced observation could be beneficial is recess. Traditionally, much of the research evaluating children's behavior on the playground involves naturalistic observations and has challenges with defining the observed behavior (Fanger, Frankel, & Hazen, 2012). Fanger and colleagues (2012) further looked at the construct of social exclusion for preschool students on the playground. In their coding scheme, they focused on child utterances. In a second study with child behavior on the playground, Ross and Horner (2009) evaluated a bully-prevention curriculum utilizing an observation system. Their data-collection emphasized more directly observed behavior without the emphasis on vocal behavior, rather focusing on physical aggression. Traditionally, recess observations are conducted with in-person observers; however, this can result in obstructed viewing angles. In an attempt to ameliorate these concerns, aerial drones could capture video content of the playground at an ideal viewing angle. This angle could improve the accuracy and reliability of observations.

With more advancement and development put into unmanned aerial vehicles (UAV), or drones, different sciences have been able to identify the benefits of using aerial drones to perform various types of observational research. Drones have been employed in wild-life research for safety, lower costs, and easier organization (Jones, Pearlstine, Percival, 2006); however yet to be utilized in school-based research. The purpose of this study is to directly compare ground-based with aerial-based video observations of student behavior on a playground. This exploratory study will look to see the reliability of behavioral outcome data for the use of drones for video observations.

In this study, a series of 30 observations of the students were conducted over approximately 12 weeks. These observations were conducted using a digital camera on a tripod and the camera associated with the aerial drone. Both cameras have the capability of recording high-definition (1080p) videos for a minimum of 20 minutes. The aerial drone will be flown at a 45-degree angle from the students playing, about 30 feet high and about five feet behind the perimeter of the play area. These parameters were chosen to preclude any students or observers from standing directly below the aerial drone. Further, observations were only conducted on days in which the wind is less than 10 miles per hour.

Each observation will be recorded and stored on an encrypted hard drive. Data coding and analysis will be completed in late Spring 2019. Trained observers will then use a systematic direct observation coding scheme to identify rates of student behavior observed. Coders will be blind to the session for coding but will be aware of the observation condition (ground vs. aerial). The behavioral outcome data (i.e., direct observations of students) will then be evaluated using visual analysis. Visual analysis is a method of data analysis in which a trained researcher visually evaluates the trend, level, and variability of a participant's behavior within and across conditions (Cooper, Heron, & Heward, 2006). This method is the gold standard for evaluating single case research, especially since inferential statistics are not a viable option with single-subject research. Further, the data will be compared directly to evaluate the agreement between the two videos recorded. To ensure the reliability of all observations, a second observer will also code all videos. This will then allow the calculation of inter-observer agreement by calculating the correlation of the two observers' findings.

*Bradley Bloomfield (University of Alabama)*

**D-2****Eye Movement/Eyetracking in the Assessment of Combat PTSD, TBI, Moral Injury, and Sexual Assault**

Best psychological assessment includes discriminating specificity in aspects of PTSD, TBI, Sexual Assault, and Moral Injury as all are important to the future of military and civilian Mental Health. Recent studies using eye tracking technology have indicated promising results of eye movement as a biomarker for identifying PTSD and TBI. We have completed two eyetracking studies, one study addressing assessment of PTSD and TBI and another addressing executive functioning among combat PTSD individuals. Our Research Team has found support for Eyetracking measurement and we have new information on advancing this research. This poster presentation reviews our two completed and future proposed studies. Our initial study used ETAMS eyetracking equipment and software. Our PTSD group, TBI group and control group had significant differences on

RBANS Total Percent as based on the compilation of the five index scores (Immediate Memory, Visuospatial/Constructional, Attention, Language, and Delayed Memory). There were also significant results, indicating that the mTBI groups performance was much lower than those participants in the PTSD and Control group(s). There was also significantly more cognitive weakness in the mTBI group. Our second study found that the discrimination of executive functioning among PTSD subjects by eyetracking was best assessed among subjects with active PTSD symptoms and that less neutral IAPS may be useful. Participants in the second study were administered the PTSD Checklist-Military Version (PCL-M), Behavior Rating Inventory of Executive Function-Adult Version (BRIEF-A), and a 3-minute eye tracking task. SR Research EyeLink 1000 Plus system, a non-invasive eye and gaze monitoring program, was used to administer the eye tracking task. The EyeLink 1000 Plus desktop mounted camera operates with an adjustable head and chin support to minimize the effects of head movements and increase tracking stabilization. The participant sits comfortably in a chair at the computer console while viewing 30 standardized images from International Affective Picture System (IAPS). Images are specially selected to ensure no explicit trauma-related stimuli (e.g., people, infants/children, and/or vehicles) are present. Neutral images (e.g., scenes of landscape and nature) were used. This was designed to minimize the likelihood of complicated patterns of negative emotion and other PTSD-related symptoms being evoked during stimuli presentation. All selected images had one or more interest areas (IA), which were not visible to the participant. The pattern of dwell time and fixation count to each IA were analyzed using SR Research EyeLink Data Viewer proprietary algorithms and statistical analysis software to distinguish between individuals with or without PTSD. Following the studies, participants received a short debriefing. This ensures uncomfortable feelings, memories, or mental health symptoms that may have emerged were addressed.

Recruitment of subjects in both studies included a review of clinical records for potentially qualified participants, flyer distribution, phone call with prior notification, and a word of mouth screening from the research team. Inclusion criteria for enrollment included being a Veteran or Active Duty service member and having at least one combat-deployment. Veterans with recent substance abuse or visual difficulties were excluded. Knowledge gained from the studies increases insight into the nature of PTSD and TBI relative to particular patterns of eye movement. Findings from our initial studies were in accord with a growing body of evidence that Veterans with combat-related PTSD and TBI had eye movement patterns that may assist in the clarity of diagnosis. Our next study will attempt to further clarify eyetracking patterns and the relationship between variables relevant to PTSD and malingering. We have additional proposals to study sexual assault victims and Moral Injury subjects with eyetracking.

*John Bair (Lovell Federal Health Care Center)*

### D-3

#### **Reaching the Latino Population: A Brief Conceptual Discussion on the Use of Telehealth to Address Healthcare Disparities for the Large and Growing Population**

Latinos collectively rank as the largest ethnic minority population in the USA, with estimates suggesting continued growth into the future. Despite the Latino population being identified as a demographic in need of greater medical and mental health care, a lack of adequate services, as well as underutilization, persists. Specifically, local availability of services, language barriers, transportation issues, health insurance coverage, immigration status, and acculturation levels are among the many challenges precluding Latinos from receiving appropriate care. Telehealth has been suggested as a novel means of addressing recognized disparities, specifically for Latinos who have been indicated as a population with greater utilization of the Internet and associated technologies than other population groups. As such, the purpose of this conceptual discussion is to briefly detail the unique possibilities that the integration of technology can hold for Latino-focused medical and mental health care. Opportunities for improved education, prompting of recommendation usage, monitoring of vitals, and application of other direct treatments are discussed in relation to the most common medical and mental health issues faced by Latinos.

*Michelle Martinez (Midwestern University)*

### D-4

#### **Virtual Hope Box Buffers Acute Distress Among a Complex Older Adult Veteran Cohort**

While the non-pharmacological management of distress among the older adult Veteran cohort with complex co-occurring medical and psychiatric conditions living in long-term care facilities is of critical importance, traditional psychotherapy interventions pose challenges (Bohart & Tallman, 1999, Hartz, Brennan, Jasdeep, & Estrin, 2010, Manepalli, Desai, & Sharma, 2009). To bolster the management of distress among the Veteran cohort within a Veterans Affairs (VA) extended care Community Living Center (CLC), an array of mobile applications were utilized with Veterans as part of their ongoing mental health care treatment plans. Among the applications, a Department of Defense/VA developed mobile application called "Virtual Hope Box" was utilized with Veterans to see if the use of the application had a positive clinical impact on Veteran distress. Clinical impact information was gathered as part of ongoing clinical process improvement efforts within the clinical environment. A sample of 26 Veterans were provided technology supported psychotherapy interventions that included the use of the Virtual Hope Box mobile application. A dependent samples T-Test was utilized to explore if the technology-enhanced intervention had a significant effect on Veteran self-reported distress. Findings show that there was a significant difference between average Veteran rated subjective distress ratings before intervention and average Veteran rated subjective distress ratings after intervention ( $t(25)=4.774, p<.001$ ). The effect size for the difference was large ( $d=.937$ ). A follow-up evaluation on a subset of those Veterans who were also evaluated by a provider rated multidimensional behavioral distress scale shows a significant reduction in average Veteran behavioral distress pre and post intervention ( $t(22)=5.586, p<.001$ ). This difference was of a large effect ( $d=1.165$ ).

These results suggest that technology assisted psychotherapy using the Virtual Hope Box application may help to reduce self-reported and behavioral distress often present among older adult Veterans with complex co-occurring medical and psychiatric conditions living in long-term care settings.

*Jonathan Sills (VA Palo Alto Health Care System)*

## D-5

### Using Technology to Advance Post-Diagnostic Support: Prototype Development and Initial Feasibility Testing

In this proposal, we aim to share our journey turning knowledge learned from the literature into a prototype that aims to advance the post-diagnostic support for children with Autism Spectrum Disorder (ASD).

**Problem Statement:** Transitioning from hearing a diagnosis of ASD into intervention is a pivotal stage in caring for a young child with ASD. Unfortunately, the literature has shown that the diagnostic and treatment systems are drastically disconnected (Gordon-Lipkin et al., 2016), which leads to treatment delay, a prolonged time period between receipt of the ASD diagnosis and receipt of treatment (Wong et al., 2016). Treatment delay is caused by a range of structural and psychological barriers including a lack of quality information and post-diagnostic support (Bitsika & Sharpley, 2004). Caregivers have reported that they feel uninformed about the diagnosis and puzzled about the next step after the diagnosis of ASD was given (Braiden et al., 2010). In the current system, after long waits and extensive diagnostic testing, caregivers of children diagnosed with ASD meet with their clinician and obtain a report about the diagnoses and recommendations (Austin, Katz, & Reyes, 2011). Ideally, caregivers will be activated and able to act on recommendations in order to obtain intervention and services. However, evidence has long shown that this face-to-face, one session feedback appointment is not effective (Gordon-Lipkin et al., 2016).

**Procedure:** Our team first gathered and analyzed data from parents whose children recently received a diagnosis of ASD and clearly delineated and conceptualized the treatment delay phase experienced by families (Wong et al., 2016). Then, partnering with a nationally recognized children's hospital, some internal data were reviewed regarding its diagnostic and post diagnostic processes. An internal study (Storer, 2014) revealed that caregivers are often confused, unable to remember all of their children's diagnoses, and not acting on treatment recommendations. Moreover, the feedback loop is broken as there are no established communication channels to route data regarding parent understanding and action back to the clinicians to prompt more individualized, timely support.

After pinpointing the problems, our team uses participatory research methods to involve families in the design and development of an Automatic Learning Assistant System which aims to bridge the diagnostic and treatment systems by using the latest technology to provide a one-stop, one-click support for parents. Essentially it turns written diagnostic reports into individualized, proactive, actionable buttons on a phone application to facilitate access to treatment and support. Features include (1) an action management platform, (2) algorithms that individualize action steps, and (3) a participatory feedback channel. The involvement of parents in the early stage (e.g., provided feedback on part of the design and roadmap of the development of ALAS) has increased the social validity of the potential effective of ALAS.

**Analyses and Results:** An initial, functional prototype of ALAS will be completed in May 2019. Twenty parent users and five psychologists will participate in the pilot testing using the updated ALAS by running 5 scenario cases (Medlock, Wixon, Terrano, Romero, & Fulton, 2002; Lewis, 1990, 1991, Medlock et al., 2002). Data with regard to feasibility, usability, and a failure analysis will be available for the conference.

Our team will also share the facilitators and barriers at the system level (e.g., HIPPA compliance, funding, patient use of technology, ownership issues, patient participation) with regard to innovation around post-diagnostic support. We will also share particular challenges in automating the post-diagnostic support ASD (e.g., accuracy of information about providers and resources, insurance).

**Implications:** By sharing our prototype development process, clinicians/ researchers will learn more about (1) ways to generate a preliminary prototype using participatory research methods and (2) gain insights into ways of turning clinical insights and research findings into a prototype.

*Venus Wong (UC Davis)*

## D-6

### Strategies to Telemental Health Technology Adoption

Telemental Health (TMH) within the Department of Veterans Affairs (VA) refers to the utilization of videoconferencing to deliver psychotherapy and pharmacotherapy services to patients in behavioral health settings. This is a cost-efficient and promising way to enhance psychological healthcare delivery to military veterans in the United States and globally. Not only has TMH technology been found to be effective in delivering psychological interventions across a wide range of diagnoses (i.e., depression, anxiety, psychoses, posttraumatic stress disorder (PTSD), substance use, and suicide prevention) among military veterans, it also does not compromise the therapeutic alliance. For instance, patients have reported high levels of satisfaction with TMH services. Additionally, when used in rural communities with a lack of resources or in locations where recruiting and retaining behavioral health clinicians is challenging, TMH could alleviate these barriers to accessing psychological health services. The use of TMH technology has also been found to be effective in the recruitment and retention of clinicians. Yet, despite its effectiveness, previous research has shown that clinicians are hesitant about the use of TMH videoconferencing technology to provide behavioral health services. As the need to provide psychological services to military veterans in the rural areas continues to increase, it is imperative that VA clinicians should explore ways to implement innovative approaches to address the demand of psychological services and move toward a long-term solution. A quality improvement (QI) survey composed of 25 questions was distributed to clinicians working in a VA Health Care System in the western United States. This survey explored clinician's perceptions of TMH use, perceived and/or real barriers to its adoption, previous formal TMH training and previous experience with TMH, and clinician's preferences for the different types of TMH modalities available. The qualitative data were analyzed and the implications related to the use of TMH were compiled. From the results of the study, the following recommendations were generated: 1. Provide TMH formal training to address barriers to the adoption of TMH technology, 2. Incorporate required TMH training for students and faculty in psychological health-related academic

programs at the university level; 3. Provide relevant diversity training to address the cultural, social, and geographic characteristics of military veterans from rural areas, 4. Develop and implement strategies to research to study this area, and promote use of TMH technology in provision of evidence-based interventions in clinical practice.

*Desiree Nguyen (San Francisco Department of Veterans Affairs)*

## D-7

### Development and Testing of a Virtual-Nurse Led Mobile Health Intervention for Patients with Chronic Conditions

**Background:** The care transition from hospital to home is among the most vulnerable periods along the care trajectory, requiring a combination of both physical health and motivational assistance. Patients with complex, chronic health conditions are at an increased risk of adverse events during this time, including re-hospitalization. Upon arrival home, patients are tasked with further self-managing their care; unfortunately, many are ill-equipped, unmotivated, or unsure how to handle their own needs throughout this experience. A growing approach to address these challenges is through patient-facing technologies that leverages relational agents—or embodied conversational agents—to present information and interact with patients in a more social, motivational way. We report the results of our multidisciplinary process for developing, testing, and refining a technology-assisted care transition intervention, founded on the concept of a virtual nurse (Nurse Annie). Our mobile intervention leverages multiple technologies, including a tablet-based relational agent and automated short message service (SMS) text messaging. The intervention purpose is to engage and educate patients about the care transition experience during the inpatient hospital stay, and to support their self-management efforts once they are discharged home.

**Methods:** This work was conducted in the Veterans Health Administration (VA). Based on our process, we present results related to conceptual development, alpha and beta testing, and implementation planning across multiple levels of feedback and assessment. A combination of quantitative data (patient feedback on usability, satisfaction, and usefulness) and qualitative feedback (interviews with patients and major stakeholders) were collected to aid in iterative development and revision.

**Analyses:** We describe the development and initial testing of our intervention across four interconnected processes. In Process I, intervention development and conceptualization, the research team consulted with psychologists, behavior change experts, physicians, software engineers, and patients regarding relational-agents, care transition interventions, and clinical management to design intervention content. Process II, alpha testing of the two intervention technologies, was conducted with veteran patients in a controlled environment to gather feedback on the virtual nurse, usability, intervention content, and overall satisfaction. In Process III, beta testing of the intervention components, veterans with relevant diagnoses tested the intervention technologies and provided usability and satisfaction information via surveys and interviews. Process IV, implementation planning, was completed through site visits and 13 interviews conducted with stakeholders at VA facilities where the intervention was scheduled for roll-out as part of a multi-site, randomized trial.

**Results:** Process I findings suggested that the content addressed through the relational agent interaction was understandable, useful, engaging, and of an appropriate length. Veterans suggested edits

to the agent and text message content, requesting more information related to medication adherence. In Process II, veterans found interactions with the relational agent to be simple, compassionate, relatable, and useful; however, some voiced concern that other veterans may be reluctant to interact with the relational agent because of their fear of technology. Following iterative development, in Phase III, a majority of Veterans (75%) described the relational agent interaction to be a positive experience, with no texting difficulties (80%). In Phase IV, stakeholders across the three hospital facilities participating in the randomized trial provided further feedback for intervention presentation and alignment with current clinical processes, and agreed with the benefit of our intervention, and emphasized the complex needs of Veterans.

**Practical Implications:** The result of this development and testing process was a highly usable and socially interactive mobile-health tool that is being tested as part of an intervention in a three-site randomized controlled trial. Based on our process, we also present a generalized development, testing, and implementation planning model that can be used to guide development of future technology-assisted interventions.

**Conclusion:** Our process drew from prior mHealth research and user-centered design to create a mobile intervention aligned with the post-discharge needs of both patients and their clinical care staff. To create a personalized and interactive technology intervention, multiple iterations of feedback—by behavioral change experts, clinical staff, and patients—was necessary to make sure the virtual-nurse led intervention was effective, relational, and technologically functional. Following initial conceptualization, iterative revisions occurred within each stage of alpha and beta-testing as well, which helped identify and create solutions for issues such as texting concerns, feedback regarding the social nature of the virtual nurse, and potential difficulties for patients with auditory or cognitive issues related to information intake.

*Erin Reilly (Edith Nourse Rogers Memorial VA Medical Center)*

## D-8

### Nightmare Deconstruction and Reprocessing: Integrating Collection of Physiologic Stress Indicators via Wearable Technology and Genomic Analyses in Pilot Testing a Psychotherapy for PTSD-Related Nightmares

**Statement of the Problem:** Nightmares and insomnia are associated with the etiology and chronicity of posttraumatic stress disorder (PTSD) and are often refractory following evidence-based treatment (EBT) for PTSD. Current EBTs, such as Prolonged Exposure (PE) and Cognitive Processing Therapy, utilize fear memory extinction and reconsolidation but do not target sleep symptoms. Given that dreams are believed to play a role in learning and memory consolidation, developing a psychotherapy that utilizes nightmare content to activate and reprocess trauma memory may facilitate reconsolidation.

Memory reconsolidation requires reactivating trauma memory, which triggers a stress response. Assessing biomarkers of stress would be an effective tool when developing treatments that utilize exposure and reconsolidation. Electrodermal activity (EDA) and heart rate variability (HRV) have been used to measure stress response during PE. In addition, wristband actigraphy is a validated technology for monitoring rest and activity as an indicator of sleep disturbance. The Empatica E4 wristband is a device that gathers HRV, EDA, and

actigraph data; however, it has not been used to assess both in-session stress indicators and sleep disturbance. Regarding genomic and inflammatory biomarkers of stress, brain-derived neurotrophic factor (BDNF) mediates the effect of stress on neuronal plasticity and is correlated with sleep disturbance, stress, and fear memory consolidation. Val66Met, a BDNF single nucleotide polymorphism has also been associated with PTSD etiology and extinction learning. PTSD, insomnia, and treatment efficacy are associated with inflammatory biomarkers such as IL-2, IL-6, and CRP. Because expression of BDNF, IL-6, and other biomarkers vary throughout the day, collecting blood samples should be done within a consistent circadian window.

The current study is testing Nightmare Deconstruction and Reprocessing (NDR), a three-stage treatment that integrates exposure and reprocessing of nightmare images to facilitate trauma memory reconsolidation. Study aims are to test NDR's plausibility and tolerability and to test the methodologic feasibility of collecting daily HRV, EDA, and actigraphy data via the E4 wristband as well as blood samples for genomic analyses at three time points within a specified circadian window. Our research questions are: 1. Regarding NDR's efficacy and tolerability, (a) What is the mean baseline-to-post-treatment change in nightmare and insomnia severity? (b) What proportion of participants dropped out? (c) How does dropout relate to HRV, EDA, and self-reported distress? 2. Regarding the research utility of E4 data collection, (a) Did participants comply with daily data download? (b) How well does E4 data show sleep disturbance and in-session HRV/EDA during exposure to nightmare images? 3. Regarding feasibility of blood sample collection, did collection conform to the 0800-1200 circadian window?

**Procedures:** Participants are combat veterans (N = 30) ages 18 to 64 with trauma-related nightmares and insomnia who are being recruited at Walter Reed National Military Medical Center. Participants undergo 8 NDR treatment visits over 8 weeks and complete a 1-month post-treatment follow-up. Psychometric measures of nightmare (Disturbing Dreams and Nightmare Severity Index; DDNSI) and insomnia (Pittsburgh Sleep Quality Index; PSQI) are administered at each treatment visit. Objective measurement of sleep disturbance is done via the E4 actigraph. Participants are instructed to download their E4 data daily. In-session stress is measured using the 1-item Subjective Units of Distress Scale (SUDS), which is administered 3 to 5 times in each session as a measure of participants' distress during exposure to each nightmare image. In addition, HRV and EDA data collected via the E4 wristband are objective indicators of in-session stress during exposure to nightmare images. All NDR sessions are audiorecorded so that HRV and EDA epochs can be synchronized with moments in session when participants are exposed to nightmare images. In addition, BDNF, IL-2, IL-6, and CRP will be assayed from blood samples taken at three points: Visit 0 (baseline), Visit 1 (first in-session exposure to nightmare images), and Visit 7 (final in-session exposure). Blood samples are collected immediately following the NDR treatment session, within a prescribed circadian window (0800-1200).

**Analyses:** Plausibility of NDR's efficacy will be determined by calculating effects sizes (Cohen's *d*) of pre- to-post-treatment change in nightmare and insomnia severity (DDNSI, PSQI, and E4 actigraph data). Treatment tolerability will be determined by calculating percentage of dropouts and by examining the relationship of dropout to in-session stress markers, including HRV, EDA, and SUDS. Participant compliance with E4 download will be determined by calculating the mean number of days of missed downloads, the

proportion of participants who comply with at least 75% of downloads, and how compliance relates to treatment outcome. Blood sample collection compliance will be determined by examining time of collection, how many samples occurred outside the 0800-to-1200 window, and what the barriers were to timely collection.

**Results:** Data collection is underway; preliminary results will be reported.

**Practical Implications:** Study results may provide evidence of treatment and methodologic feasibility in order to move forward with a large-scale randomized clinical trial.

*Patricia Spangler (Uniformed Services University)*

## D-9

### Exploring Market Readiness for a Digital Mental Health Application to Enhance Resilience

**Background and Rationale:** The World Health Organization estimates that mental health affects one in four people worldwide. Although options for treatment are available, nearly two-thirds of those currently suffering from mental health concerns encounter barriers to seeking treatment (Andrade et al., 2014). These barriers include stigma associated with mental illness, lack of accessibility to services, and the high costs associated with obtaining and maintaining treatment. Further, focus on mental health has shifted from treatment to prevention, as our collective understanding of mental health has graduated from a false dichotomy of sickness and health to a continuum of mental experience (Galderisi, Heinz, Kastrup, Beezhold, & Sartorius, 2015; Keyes, 2002; World Health Organization, 2004). Taken together, the barriers to treatment and the emphasis on prevention suggest that mental health approaches could be designed both to increase access and to promote holistic, preventative care. Finally, the digitization of services via e-health/telehealth applications have been suggested as a useful approach in tackling existing limitations in mental health (Anthes, 2016). Yet, the extent to which the general public is accepting of these shifts in mental health management and care provision is unclear.

**Procedure:** The current research aims to explore market readiness for digital mental health applications across Canada, the United States, and the United Kingdom. Using Facebook Campaign test and Amazon Mechanical Turk, we sampled over 150 participants across three countries on their readiness to engage with a digital mental health application aimed at promoting resilience and wellbeing.

**Results:** Qualitative and quantitative analyses both converge to suggest that a large majority would feel comfortable utilizing a digital application to engage with their mental health. Further, approximately half of the respondents surveyed indicated their willingness to pay for this application. We then piloted an evidence-based, holistic mental health curriculum in a small sample of full-time executives (N = 13). Preliminary results suggest that participants enjoyed the flexibility and insights gained from an online digital application that offers anytime access to a full spectrum of mental health content. Pilot data collection for pre- and post-digital application intervention is ongoing, with an expected completion date to be June 2019. Pre- and post-intervention changes across self-reported measures of resilience will be examined to determine the efficacy of online e-interventions promoting mental health.

**Implications and Conclusions:** Although research has identified a gap in the delivery of process-based preventative mental health care

in a way that is accessible to the individual, the availability or quality of real-life applications have yet to meet those needs. The current research evaluates the readiness of individuals across three countries to engage with consumer-grade digital applications aiming to promote mental health and resilience, as well as provide preliminary evidence for its efficacy in a small sample of participants. Findings paint a compelling possibility for the future of mental health service delivery to be incorporated with technology through the digitization of evidence-based practices via e-health and tele-health services.

*Jenny Jing Wen Liu (Ryerson University)*

#### D-10

##### Evaluating Psychotherapist's Use and Attitudes towards Mobile Applications

Mobile health (mHealth) is an emerging area providing mobile technology users with mobile applications intended to assist individuals with a variety of goals, such as behavioral and emotional change as well as symptom management and monitoring. The benefits of using mHealth, such as accessibility and low cost, has contributed to the rapid development of mobile applications for psychological and health concerns. In addition, the Pew Research Center (2012) previously found that one fifth of all smartphone users reported using a health-related mobile application.

Using mHealth as an adjunct to psychotherapy has been found to be both feasible and have the potential to enhance face-to-face treatment (Levin et al. 2017). However, information on the attitudes of psychologists and clients towards using mHealth is limited. Specifically, the proportion of psychologists who use mental health applications in their practice is unclear and whether psychologists may choose to use mHealth in their practice is unknown. Shedding light on these questions will help to further understand the use of mobile applications in mental health treatment.

In order to explore these questions, researchers aim to recruit a nationally representative sample of licensed psychologists. Currently, recruitment of psychologists from New Jersey is complete (N=79). Researchers plan to continue recruiting psychologists from 12 additional states across the United States. The sample is currently 62% female, with an age range of 28 to 81 (M=55.29; SD=13.39). In addition, 87% of participants identified as White or Caucasian, 65% of participants indicated to work in private practice, and 61% of participants reported they have been licensed for at least 15 years. Participants completed an online survey of use and attitudes toward mhealth applications for mental health issues.

Participants were asked to rate how much they agreed with statements pertaining to their views of using mobile applications clinically on a 5-point Likert scale, ranging from (1) Strongly Disagree to (5) Strongly Agree. Results indicated that 71% of psychologists indicated that they at least somewhat agree that mobile applications could be useful in helping their clients overcome mental health concerns. In addition, 52% of psychologists either agree or strongly agree that they would like to further incorporate mobile applications into their practice in the future. However, psychologists also reported a number of apparent barriers to using mobile applications as an adjunct to treatment. Specifically, 65% of psychologists reported no familiarity with research literature on mobile applications, and 84% of psychologists reported being unaware of any resources that could be used to make informed decisions regarding application utilization.

Finally, when asked what concerns psychologists might have about implementing applications into their practice, the most commonly reported reason was that they felt unsure of whether they had the knowledge to make effective use of different mobile applications.

Overall, results suggest that psychologists tend to agree that clients would benefit from using mobile applications and that they would like to further incorporate the use of mhealth into their practice. However, results also highlight important barriers that need to be addressed in order for mobile apps to have a significant impact on mental health practice. Of particular importance might be the development of resources to help guide psychologists who do have an interest in integrating mobile apps into their clinical practice.

*Sean Martin (Rowan University)*

#### D-11

##### Older Adults' Preferences Towards Modality and Location of Cognitive Testing: Implications for the Feasibility of Teleneuropsychology

The importance of early recognition of Mild Cognitive Impairment (MCI) and dementia among older adults is widely recognized by geriatric experts and stakeholder organizations. Alzheimer's disease, the most common form of dementia, is characterized by a slow, gradual, and insidious prodrome. The diagnostic process of MCI and dementia involves extensive cognitive testing. Timely detection and intervention have shown to lead to better health outcomes and reduced healthcare costs (Getsios et al., 2010; Barnett et al. 2014). Barriers to cognitive assessment including cost and mobility issues can delay crucial detection and, ultimately, delay intervention. Innovations in technology can ameliorate this issue by allowing for convenient and cost-effective cognitive testing. In the current study, we examined the preferences of community-dwelling older adults towards traditional and novel test platforms.

Applications of teleneuropsychological testing are gaining momentum in providing a feasible alternative to clinic-based assessment. Recent applications of teleneuropsychological testing include telephone interviews, videoconferencing, virtual-reality tests, and self-administered tests (Munro Cullum et al. 2014; Tierney et al. 2014; Parsons et al. 2017; Plancher et al., 2012; Kennedy et al., 2014). These applications may help address the reluctance to engage in cognitive testing and the unfavorable testing experiences reported by many adults. For example, individuals who complete cognitive testing are more likely to highlight their failures when assessing their performance (Krohne et al. 2011). Consequently, the testing experience elicits feelings of shame and irritation (Krohne et al. 2011).

The need for less distressing testing experiences calls for person-centered care in the cognitive assessment context. Person-centered care involves incorporating and respecting the patient's values, needs, and preferences in all clinical decisions (Epstein et al., 2010; Fraenkel, 2013). Providing person-centered care has shown to improve quality of life and disease outcomes in patients. It has also played a critical role in addressing disparities in racial, ethnic, and socioeconomic health outcomes (Epstein et al., 2010). Research has also shown that person-centered care improves patient satisfaction and quality of care while reducing health care costs (Epstein et al., 2010). Both clinic-based assessment and remote assessment approaches require proper investigation into patient's preferences surrounding the testing experience in order to take steps towards person-centered care.

This study is the first of its kind to examine older adults' preferences towards the modality and location of cognitive testing. We wanted to determine the percentage of individuals who preferred their home as a testing location, the computer as a testing device, and the combination of home-based computerized testing. Adults 50 years or older completed the Attitudes Around Cognitive Testing questionnaire (Wong and Jacova, 2018) at primary care sites and through an online crowdsourcing site ([www.mturk.com](http://www.mturk.com)). AACT examines respondents' preferences regarding the modality (computer, paper/pencil, mobile device, and/or no preference) and location (home, doctor's office, and/or no preference) of cognitive testing. This sample included 409 respondents (300 online, 109 in primary care), with mean age 62.8 years (range 50-91), the majority being females (65.2%). Among respondents, 28.4% preferred home-based testing, 12.2% preferred testing in a clinic, and 5% reported multiple preferences. Over half (58.2%) reported having no preference. As for modality, 19.3% preferred computerized testing, 5.4% preferred paper and pencil testing, and 18.3% reported multiple preferences. Similar to location, over half of our respondents (56.5%) reported having no preferences for modality of testing. When only respondents preferring home-based testing were considered, 36% preferred computerized testing, 6.9% preferred paper and pencil, 33.6% had no preference, and 23.3% had multiple preferences for the modality of testing. Those preferring home-based computerized testing attached substantially more importance to this choice on a 5-point scale than those with no/other preferences (4.2+0.9 vs. 3.3+1.2,  $p=.000$ ). Chi-square revealed that a higher proportion of online respondents preferred testing at home, compared to respondents in primary care (33.6 vs. 13.8%,  $p=.003$ , 30.5 vs. 8.4%,  $p=.000$ , 12.7 vs. 3.7%,  $p=.008$ , respectively). We did not identify associations between demographic and clinical variables including age, gender, education, subjective cognitive decline, and preferences for home, computer, or home-based computerized testing.

Our findings suggest that a sizable proportion of older respondents would choose the home environment for cognitive testing. This finding highlights the opportunities that exist for teleneuropsychology application. However, only a small percentage would choose the computer as a testing device in the home setting. It is clear that such a preference is more prevalent among older adults familiar with the computer (our online respondents). We therefore conclude that, at least for now, teleneuropsychology may need to consider multiple platforms including the telephone rather than only computer technology. This may change when the older adult population includes cohorts familiar with computer technology. These findings are novel and encourage the integration of person-centered considerations and novel test platforms into standard neuropsychological assessment.

*Moriah Splonskowski (Pacific University)*

#### D-12

##### **Real-World User Engagement Characteristics of a Globally Available Depression Self-Help Smartphone App**

mHealth (mobile health) interventions for mental health have shown promise in feasibility, usability, and efficacy (Cuijpers et al., 2010; Donker et al., 2013; Firth et al., 2017). However, a majority of studies are limited in sample size, and versions of the researched interventions are rarely available to the public. Very few studies evaluate mobile mental health technologies that are available in real-world

settings with real-world users. Consequently, little information is known about how individuals engage with these interventions in ecologically valid ways. The present study examines user behavior patterns in a self-help smartphone application for depression called MoodTools. MoodTools contains six features grounded in evidence-based principles for the treatment of depression: psychoeducation about depression, self-monitoring of depressive symptoms (PHQ-9; Kroenke, Spitzer, & Williams, 2001), thought diary, strategies for behavioral activation, development of a suicide safety plan, and links to additional resources. This study evaluates how those who download MoodTools use it.

Anonymous, aggregate user data was collected via Google Analytics. We evaluated MoodTools usage from a cohort of  $N = 33,993$  users who downloaded the Android version of the app from June 10, 2018 to February 28, 2019. We examined age group (18-24, 25-34, 35-44, 45-54, 55-64, 65+), gender, country of origin of download, week-to-week user retention rates, and average time spent in app. User retention was measured by segmenting users into weekly cohorts to analyze those who return to the app over a span of eleven weeks.

Users downloaded MoodTools across the Americas (49.92%), Europe (27.15%), Asia (14.24%), Oceania (5.03%), and Africa (3.55%). The app has been downloaded in 160 countries, with top countries being the United States (39.27%), United Kingdom (12.85%), Canada (5.93%), India (5.61%), and Australia (4.06%). Women downloaded the app more than men by a ratio of 2:1, while two-thirds of users were between ages 18 and 34. Women downloaded the app more than men in top countries of download with the exception of India. On average, 17.9% of users who downloaded the app at week one returned during week two, 11.1% during week three, and 7.85% during week four. By week twelve, an average of 2.81% users were still using the app. The average length of time spent using MoodTools was 5 minutes 8 seconds, and there were no practical differences in amount of time spent in app by age group.

These findings offer more ecologically valid understandings of user behavior of a publicly available smartphone app for depression and may also help inform future app development. Although MoodTools is only offered in English, worldwide download numbers suggest a global need for mHealth interventions for mental health. Drop in user retention rates after the first week point to the importance of implementing engagement strategies to maintain interest over time.

*Langting Su (Georgia State University)*

#### D-13

##### **The Effectiveness of an App-Based Cognitive Behavioral Therapy Program for Alleviating Sleep Disturbances Among Cancer Patients**

Cancer patients experience psychological and social difficulties (e.g., depression, anxiety, fatigue, & pain starting from the time of diagnosis throughout the survival period (Min et al, 2013; Greer, 2002; DeJong, & Fombonne, 2006). 30% to 60% of cancer patients are reported to have sleep disturbance, including insomnia, excessive sleepiness and difficulty breathing while sleeping (Davidson et al 2002; Le Guen et al., 2007; Garland et la., 2014). This is significantly higher than the reported rate of 9-12% of all adults who experience chronic sleep disturbances (Ohayon & Hong, 2002; Riemannental,

2011). Sleep disturbances are associated with fatigue and depression (Donovan & Jacobsen, 2007), and in addition, lower the threshold of pain (Satia et al., 2008). Furthermore, sleep disturbance impairs immune cells (Patel et al., 2009), leading to changes in carbohydrate metabolism and endocrine function (Spiegel et al., 1999), which can slow the recovery of cancer patients and worsen the disease (Blask, 2009). However, despite the high prevalence and serious effects of sleep disturbances, patients and clinicians undervalue the need for treatment, and often do not receive adequate diagnosis and treatment (Mystakidou et al., 2007). Effective evidence-based treatments for sleep disturbance in cancer patients include cognitive behavioral therapy, exercise and medication (Howell et al., 2013). Among them, cognitive behavioral therapy (CBT) is a psychological intervention that focuses on changing emotions, behavior, and distorted perceptions, or learning coping skills for problem solving (Beck, 1993; Ellis, 1962). However, CBT is typically delivered individually via face-to-face format by professionals (Mohr et al., 2012), and requires lots of time and economic costs. (Beatty et al., 2016).

Recently, mobile health, which uses mobile technologies such as smartphones, tablets and wearable devices to provide physical and mental health-related interventions (Lui, Marcus, & Barry, 2017), has gained attention as an alternative method, mainly because it can provide necessary services beyond the constraints of time and space (Smith et al., 2016; Watts et al., 2013). Most of the mobile app-based CBT programs and studies that have been developed are related to depression and anxiety, and not sleep disturbance. Up until now, only one study was conducted using an app-based CBT program for sleep disturbance (Horsch et al., 2017). The study investigated the effectiveness of an app-based CBT program with 151 participants who had sleep disturbances. The results showed a significant decrease in sleep disturbance symptoms in the training group with app-based CBT, compared to the control group. The purpose of this study was to develop an app-based CBT program for sleep disturbance in cancer patients (HARU Sleep) and to investigate the effects of this program on the levels of sleep quality using subjective and objective measurement tools. For this purpose, 37 cancer patients with sleeping disturbances were recruited via referrals from the 3 major hospitals in the Seoul metropolitan area, and through internet advertisements. 4 participants who were taking sleep medication were excluded from the study. Participants were assigned to one of the following 3 groups: 1) a CBT program, 2) a simple, information-providing program, or 3) a waitlist control group. All participants completed the following assessment instruments before and after the training: Pittsburgh Sleep Quality Index (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989; 최희정, 김성재, 김범중, & 김인자, 2012), Dysfunctional Attitude Scale (Weissman & Beck, 1989; 권석만, 1994), Health-Relation Quality of Life scale (Ware & Sherbourne, 1992; 고상백, 장세진, 강명근, 차봉석 & 박종구, 1997)) and 2 computer tasks consisting of a dot-probe task (MacLeod, Mathews & Tata, 1986) and an implicit association task (Greenwald, McGhee, & Schwartz, 1998).

To test the effects of the CBT program, multilevel analyses were conducted. Models were built to explore within-group (time), between-group (conditions: CBT group vs simple information group vs waitlist group), and interaction (time X condition) effects. The results showed that the sleep disturbance symptoms were significantly decreased in the group that used the app-based CBT program compared to the waitlist group. Furthermore, health-related quality of life scores were significantly increased in the CBT group compared to the waitlist group.

On the other hand, there were no significant differences in the dysfunctional attitude scale and the computer tasks in any of the groups. Such results suggest that app-based cognitive behavior therapy programs may be effective as interventions for the alleviation of sleep disturbance in cancer patients. The limitations of this study and suggestions for future research are discussed.

*Changseok Lee (Yonsei University)*

## D-14

### Examining Evidence-Based Content in Smartphone Apps for Depression and Anxiety: A Review of Publicly Available Apps

**INTRODUCTION:** Most people in need of mental health services do not receive them (Kazdin & Blaze, 2011), but technology-based interventions have the potential to expand access to care (Fairburn & Patel, 2017; Kazdin, 2017). Smartphone applications may be especially potent; approximately 65% of smartphone users report using an app for mental health symptoms in the past month (Rideout & Fox, 2018). Although some smartphone applications for mental health treatment (MH apps) have empirical support (Firth et al., 2017; Josephine, Josefine, Phillip, David, & Baumeister, 2017), most empirically supported interventions are not publicly available (Fleming et al., 2016). Even those some popular publicly available MH apps have acquired millions of downloads and hundreds of thousands of active users (e.g., Headspace), relatively little is known about the content within popular MH apps. In this study, we applied the distillation and matching model (Chorpita, Daleiden, & Weisz, 2005) to popular MH apps for depression and anxiety in order to evaluate the presence or absence of evidence-based content.

**METHOD:** In July of 2018, we searched the Apple App Store and Google Play Store using the terms "depression", "anxiety", and recommended search terms which were automatically generated by each app store ("Depression Games", "Depression Helper", "Depression Tracker", "Depression Wallpapers", "Depression and Anxiety", "Anxiety Relief", "Anxiety Relief Apps", "Anxiety Relief Games", "Anxietyhelper", and "Anxiety & Panic Attacks"). Because app users very rarely scroll past the first five apps (Dogruel, Joeckel, & Bowman, 2014), we limited our search to the first five apps from the search results for each search term. We included apps that offered some form of treatment or support to individuals; we excluded irrelevant apps (e.g., apps that provided phone screen backgrounds or inspirational quotes). The first author screened apps for inclusion; the second author reviewed the titles and descriptions of all excluded apps.

The distillation and matching model (Chorpita, Daleiden, & Weisz, 2005) involves identifying "common elements" (Chorpita, Becker, & Daleiden, 2007, p. 647) in psychological interventions. Common elements refer to treatment techniques that are commonly used in empirically supported treatment protocols. For example, cognitive restructuring and problem solving are common elements within empirically supported treatment protocols for youth depression (Chorpita & Daleiden, 2009). In addition to coding for common elements identified by previous studies (Chorpita & Daleiden, 2009; Higa-McMillan, Francis, Rith-Najarian, & Chorpita, 2016; Lindsey et al., 2014), we also reviewed meta-analyses (Chambless & Hollon, 1998; Chambless & Ollendick, 2001; Cuijpers et al., 2008; Cuijpers et al., 2013) and treatment manuals (e.g., Barlow et al., 2010; Strosahl &

Robinson, 2017; Weissman, Markowitz, & Klerman, 2017) to identify additional common elements.

The final codebook consisted of 25 common elements. The first and second authors independently coded one-third of apps ( $n = 9$ ) to assess inter-rater reliability, and the first author coded the remaining apps. Cohen's kappa ranged from 0.73 to 1.0 for treatment element codes (Mean = 0.97, SD = 0.08).

**RESULTS:** Our final sample consisted of 27 apps. Most apps included a few common elements of empirically supported psychotherapies (Mean = 3.70, SD = 3.52, Median = 3). Among the 16 apps for depression, the most common elements were psychoeducation (50% of apps), relaxation (38%), cognitive restructuring (31%), and activity scheduling (31%). Only 6% of depression apps included problem solving. Among the 17 apps for anxiety, the most common elements were psychoeducation (53%), relaxation (47%), and self-monitoring (24%). Only 12% of anxiety apps included exposure, 12% included cognitive restructuring, and zero included problem solving.

**DISCUSSION:** We identified common elements present in popular MH apps for depression and anxiety. We found that most MH apps include at least one evidence-based treatment elements, and some included a wide variety. Notably, we identified gaps between empirically supported treatment protocols and existing MH apps. Specifically, empirically supported treatments for depression often include problem solving (Chorpita & Daleiden, 2009), yet few apps for depression taught problem solving. Additionally, empirically supported treatments for anxiety often include exposure, cognitive restructuring, and problem solving (Higa-McMillan et al., 2016), yet few apps for anxiety included these elements.

Our findings offer several practical directions for future research. For instance, researchers and app developers may wish to develop novel MH apps to fill the gaps between empirically supported treatments and MH apps. For instance, novel anxiety apps teaching exposure, problem solving, and cognitive restructuring may be especially useful. Additionally, empirical research could examine the efficacy of certain treatment elements when delivered via smartphone applications. Certain elements may be especially difficult to effectively teach via MH apps, especially MH apps without human support.

*Akash Wasil (Harvard University)*

## D-15

### A Tripartite Model of Technology Utilization in Psychodynamic Psychotherapy

Advances in digital communication prompt revisiting psychodynamic psychotherapy, as technology can enhance and alter both the delivery of treatment, as well as the content of sessions. A continuum for the application of digital technology in psychodynamic psychotherapy is presented, consisting of three positions: tech free, tech adjunctive, and tech primary. Use of mobile applications, video conferencing, a shared online platform, or biometric devices all have the opportunity to augment sessions, which in turn may alter the therapeutic relationship between therapist and client. Given the ubiquity and persistence of computing, practitioners are encouraged to make intentional choices about their decision to use, or to not use, technology in session. Case vignettes are provided.

*Adam Freed (Madigan Army Medical Center)*

## D-16

### Virtually Healthy: Relating Virtual Reality to the Translational Research Continuum

**Objective:** From its early applications in 3D modeling and entertainment, virtual reality (VR) has found a host of uses in the modern health system, especially as health care embraces the role of technology in diagnostic, preventive and therapeutic settings. This paper intends to elaborate on VR's medical potential by discussing established clinical and health-oriented use cases, while establishing parallels between its technological adoption and the five stages of translational research.

**Method:** Following a brief introduction to the development of VR in health care and its noteworthy capabilities within the field, the framework of translational research is examined in terms of its five primary stages. The discussion then centers on existing and potential uses for VR throughout the translational continuum, assessing the technology and content design requirements for each stage. The discussion also identifies present shortcomings that may stand in the way of adoption at particular translational stages, and comments on the methodological challenges in assessing stage-related outcomes.

**Conclusion:** While further research and the maturing of VR technology are necessary to confirm its areas of greatest utility in health care, VR applications currently play a role in every stage of translational research. The unique assets of virtual worlds are promising for continued growth of VR in all stages of health care provision and research.

*Fabian Bock (University of Southern California)*

## Human-Technology / Computer / Robot Interaction

### E-1

#### Assessing the Categorization-Individuation Spectrum with Human and Robot Agents

When interacting with groups of nonhuman agents, we usually perceive them as a homogenous group of agents that have similar characteristics and capabilities (e.g., Keller & Rice, 2010). A similar effect, known as outgroup homogeneity (Hugenberg et al., 2010), is seen in human-human interactions such that social ingroup members (e.g., female faces when observer is female) are easier to distinguish perceptually than social outgroup members (e.g., male faces when observer is female). This overgeneralization is often due to a lack of experience with the other agent groups and/or insufficient motivation to see them as individuals (Geels-Blair et al., 2013). In human-robot interaction, the miscalibrated perception of individual characteristics and abilities can negatively affect trust and performance in human-robot teams, in a sense that knowledge about one robot could be overgeneralized to the remaining robot team members. The consequence is that trust in robot team members is either too high (i.e., positivity bias, if initial experiences were positive) or too low (i.e., negativity bias, if initial experiences were negative), thus resulting in overreliance on team members in the former and mistrust in the latter case (Parasuraman & Riley, 1997). This phenomenon, known as System-Wide Trust, leads to a decrease in team performance when over-relying on robots that perform below group expectations, and an increase in operator workload when under-relying on

robots performing above group expectations (Keller & Rice, 2010). A better calibration of trust could potentially be achieved if operators perceived robotic team members as individuals (i.e., individuation) that each have specific characteristics rather than a homogenous group of agents with similar characteristics (i.e., Component-Specific Trust; Keller & Rice, 2010).

Research on social perception has shown that in order to perceive other group members as individuals with specific features and abilities, our perception of them needs to shift from categorization to individuation (Hugenberg et al., 2010). Previous research suggests that individuation can be fostered by increasing observers' perceptual experience with a class of stimuli and / or enhancing their motivation to see the agents as individual beings (Fiske & Neuberg, 1990). The effect of perceptual experience on individuation was first studied with unfamiliar stimuli, the so-called Greebles, that possess certain features of human faces but are distinct enough from human faces that their individuation had to be obtained via learning (Gauthier & Tarr, 1997). Participants were exposed to families of Greebles, where all members shared certain features (e.g., long "ears"), but differed from each other on a range of other features (e.g., some had distinctively long "noses", some had very round bodies). During individuation, all Greeble family members were presented in randomized order over dozens of trials and participants were asked to learn to discriminate among them by associating each Greeble with an identifier (A to E). It was shown that as perceptual experience with the Greeble families increased, individuation abilities and activation in brain areas associated with human face perception also increased.

In the current study (109 participants on MTurk), we use a modified version of the Greeble paradigm to examine whether participants can individuate social robots and, if so, how individuation of robot agents compares to human agents. The experiment consisted of two parts: perceptual learning and recognition. Participants were presented with five human racial ingroup faces (e.g., white for white participants), five human racial outgroup faces (e.g., black for white participants) or five robot faces (i.e., Maki robot that varied in terms of eye color and shape, nostril shape, nose bridge and distance between nose and mouth); see Figure 1A. During learning, all stimuli were paired with an identifier (A, B, C, D or E) and participants were asked to learn which image belonged to which identifier. In the subsequent recognition task, participants were asked to assign the correct identifier to the displayed agent images (by selecting the correct identifier from multiple answers provided). Participants received feedback on each trial whether they correctly identified the agent or not; perceptual discrimination sensitivity was assessed over time via correct identifications (in %).

The results show that (a) individuation of robot faces is worse than individuation of human faces, (b) individuation of human racial outgroup faces is worse than individuation of human racial ingroup faces, and (c) individuation is slower for outgroup (human racial outgroup and robots) than ingroup faces (human racial ingroup); see Figure 1B. These results show that robot faces can be individuated at a rate that is comparable to the individuation of human racial outgroup faces, but that individuation performance is worse for robot than human faces in general, potentially due to decreased experience with and/or increased perceptual homogeneity of humanoid robot faces.

*Eva Wiese (George Mason University)*

## E-2

### Does Familiarity with a Social Robot Influence Outcome Valuation as Indexed by a Neural Measure of Reward in Human-Robot Interaction?

As a social species, we rely on positive and negative feedback from social contexts to calibrate our actions to reach desired outcomes (Adolphs, 2003), which includes reward-monitoring when carrying out tasks (Luck & Kappenman, 2011). For example, when engaging in gambling games in contexts where strangers are recipients of winning outcomes, individuals experience attenuation of reward-monitoring compared to when gambling for themselves as they assign less value to the winning outcome (Hassall, Silver, Turk, & Krigolson, 2016). Similarly, reward monitoring when gambling for self is enhanced in the presence of social ingroup members as opposed to social outgroup members since the motivation to win is higher when ingroup members are present (i.e., outcome processing is more efficient in the presence of an ingroup; Hobson & Inzlicht, 2016). These context-dependent differences in reward monitoring of outcomes can be indexed by a fronto-central event-related potential (ERP)—the reward positivity (RewP)—which is generated over medial-frontal brain areas implicated in action monitoring and behavioral adjustment based on reward valuation (Botvinick, Cohen, & Carter, 2004), and whose amplitude is enlarged when receiving feedback indicating a positive outcome (e.g., Holroyd, Krigolson, & Lee, 2011; Holroyd, Pakzad-Vaezi, & Krigolson, 2008). While RewP in response to gambling outcomes seems to be influenced by the degree of familiarity between the player and the recipient of the reward in human-human dyads, little is known about how RewP is influenced in social contexts where humans interact and familiarize themselves with social robots. In the current study, 36 participants performed 36 blocks (20 trials each) of a gambling task that required them to select one of two differently colored squares, one of which had a higher chance of winning (i.e., 60%) than the other one (i.e., 10%). Upon selection of one of the squares, participants received either positive ("Win") or negative ("Loss") feedback, informing them whether their selection led them to win the given trial. Which color was associated with a higher chance of winning changed between blocks so that participants had to learn the reward probabilities of each square on a block-by-block basis. Participants gambled either for themselves (self-outcome) or a social robot named Cozmo (other-outcome), with whom they either familiarized themselves through interaction (bonding) or did not familiarize themselves (no bonding) prior to the gambling task. We expected reward valuation, as indexed by RewP amplitudes, to be greater for self- versus other-outcomes independent of social bonding (in line with Hassall et al., 2016), and to be greater in the bonding versus no bonding conditions in response to other-outcomes; we also expected learning of reward probabilities of the colored squares to be faster in the bonding versus no bonding condition. Contrary to our hypothesis, however, RewP amplitudes did not differ between self- and other outcomes and were not enhanced for bonding versus no bonding for other outcomes; instead, RewP amplitude was larger for subjects in the bonding condition compared to the no bonding condition, while learning was slower in the bonding condition versus no bonding condition independent of whether the outcome affected the self or the other. This suggests that RewP amplitudes were generally larger for subjects experiencing unexpected outcomes, which is consistent with their slower behavioral learning performance. The results indicate that the degree of social bonding, or familiarity, with a robot

did not influence reward monitoring for self compared to the robot Cozmo. However, social bonding did influence RewP regardless of recipient of the outcome (i.e., Self vs. Cozmo), such that RewP amplitudes were enlarged when participants socially bonded with Cozmo. This general increase in RewPs in the social bonding condition may be due to a higher surprise value associated with positive feedback given that learning rates were slower after participants had interacted with Cozmo and thus positive outcomes were less expected (see Holroyd, Krigolson, & Lee, 2011). The absence of differences in learning between self-outcomes and other-outcomes in the bonding group suggests that there was no prioritization in reward monitoring of self over the robot. This finding suggests that interactions with social agents such as robots can influence how we behave on their behalf. Specifically, it appears that placing equal emphasis on both self and other can negatively impact learning, possibly because an enriched social environment may distract participants from the task (Umehoto, Inzlicht, & Holroyd, 2018). These findings provide important considerations of including social robots in contexts where human interaction partners must stay focused on the task at hand and perform well.

*Abdulaziz Abubshair (George Mason University)*

### E-3

#### Process Use Mediates the Relationship Between Threat-Related Reassurance Seeking and Problematic Smartphone Use

Problematic smartphone use (PSU) is conceptualized as excessive use of a smartphone that results in social or occupational functional impairment, which includes symptoms that are commonly seen in addictive behaviors, such as dependence, withdrawal, and tolerance (Billieux, Maurage, Lopez-Fernandez, Kuss, & Griffiths, 2015). PSU is related to mental health difficulties such as depression and anxiety (Elhai, Dvorak, Levine, & Hall, 2017). Both Kardefelt-Winther's (2014) Compensatory Internet Use Theory and the Interaction of Person-Affect-Cognition-Execution (I-PACE; Figure 1) model of specific internet use disorders (Brand, Young, Laier, Wölfling, & Potenza, 2016; Kardefelt-Winther, 2014) propose that individuals engage in smartphone use as a way to regulate negative emotions. As such, recent research has focused on investigating various emotion regulation strategies that may influence smartphone use and lead to PSU. Some evidence has been found that emotion regulation, in general, predicts smartphone use (Elhai, Tiamiyu, et al., 2018), and more specifically, emotional distress tolerance and mindfulness have been found to play a role in the relationships between psychopathology and PSU (Elhai, Levine, O'Brien, & Armour, 2018). To date, there have been no studies investigating associations between reassurance seeking and PSU. In addition to emotion regulation, research has investigated how types of smartphone use relate to PSU with some findings indicating that non-social (or "process") features were related to PSU (Elhai, Levine, Dvorak, & Hall, 2017). For the present study, we explored relations between threat-related reassurance seeking (a form of emotion regulation) with PSU; we also examined process smartphone use and social smartphone use as mediating variables. To clarify, social use includes the use of any feature for social engagement. Process use includes engagement in activities with non-social features and includes use for relaxation, entertainment, or news purposes. We used the emotion regulation as the predictor variable. Undergraduate students (N=296) from a Midwestern university were recruited from

a pool of introductory psychology students using the department's Sona Systems website. Participants completed a web survey, including measures of demographics, a measure of reassurance seeking (Threat-Related Reassurance Seeking Scale), a measure of reasons for engaging in smartphone use (Process and Social Smartphone Use Scale), and a measure of PSU (Smartphone Addiction Scale). Structural equation modeling was conducted using Mplus 7.4 software. In the proposed model, threat-related reassurance seeking was specified as the predictor, process and social smartphone use were identified as mediators, and PSU was specified as the outcome variable (Figure 2). Because PSU is associated with younger age (Lu et al., 2011; van Deursen, Bolle, Hegner, & Kommers, 2015) and female sex (Jeong, Kim, Yum, & Hwang, 2016; Wang, Wang, Gaskin, & Wang, 2015), age and sex served as covariates in the following analyses. First, we conducted a higher-order confirmatory factor analysis of PSU, which yielded nearly adequate fit, robust  $\chi^2(489, N=295) = 1577.59, p < 0.001, CFI = 0.89, TLI = 0.882, RMSEA = 0.087$  (90% CI = 0.084 to 0.092). Then, adjusting for age and gender, we conducted the structural equation model, which also yielded nearly adequate fit, robust  $\chi^2(962, N=295) = 1790.89, p < 0.001, CFI = 0.89, TLI = 0.88, RMSEA = 0.074$  (90% CI = 0.07 to 0.078). Process use was related to PSU,  $\beta = 0.54, SE = 0.04, p < 0.001$ , as was social use,  $\beta = 0.54, SE = 0.04, p < 0.001$ . Of these, only process use mediated the relationship between threat-related reassurance seeking and PSU,  $\beta = 0.12, SE = 0.03, p < 0.001$ . These results suggest individuals may engage in smartphone use in order to regulate their emotions by way of reassurance seeking, and they seek this reassurance via process use rather than social use.

*Abigail E. Dempsey (University of Toledo)*

### E-4

#### Robot Shares Stress: Examining the Effect of Robot Self-Disclosures in Stress-Sharing Activity

Introduction: Self-disclosing personal stress is an effective way of managing stress (Hofmann, 2012). Previous research has shown that high school students disclose more when prompted by robot that discloses vulnerability (Martelaro, et al. 2016). Thus, disclosure reciprocity in human-robot interaction (HRI) provides a promising way of offering stress intervention for both adults and adolescents (Rose and Björling, 2017). In the current study, we utilized a mixed methods, quasi-experimental design to explore how different forms of disclosure from a social robot (technical, by-proxy, and emotional) affect adult participants' attitudes and reciprocal disclosure towards the robot.

Using a between-subject design, 36 adults (19 females, M age = 21.6) were randomly assigned to converse and engage in a stress-disclosing activity with a robot that shared technical facts (technical condition), feelings from other users (by-proxy condition), or feelings from itself (emotional condition). We hypothesize that the emotional condition will elicit longer and deeper participant disclosure, and higher robot attributional ratings (such as likeability, perceived safety, user satisfaction, and intention for future use) than the technical and by-proxy conditions. Additionally, we are currently developing a qualitative coding scheme to analyze participant disclosure to the robot and their feedback during the post-interaction interview.

Procedures: Participants first completed an intake survey that captured demographics, negative attitudes toward robots (NARS),

shyness, and perceived stress. Participants then interacted with the robot alone. Using a Wizard of Oz method, the robot was teleoperated by a hidden researcher to facilitate a short conversation (Dahlbäck et al., 1993). The robot spoke a set of self-disclosure statements, each followed by a question for the participant to answer (See documentation for script). Upon completion, the participant to fill out a post-survey which included 5-point Likert scales, such as user satisfaction (4-items, like “I feel absorbed in the conversation with EMAR”), intention for future use (4-items, like “I will use EMAR again”), robot likeability (5-items) and robot perceived safety (2-items). Participants were then given a manipulation check with the question “Which of the following best describe EMAR’s style of communication?” with options: robot tends to talk about “its own emotion”, “the experience of others”, or “technical information about its system and programs”. Then, participants were interviewed by the researcher about their interaction with the robot.

**Analysis:** Degree of self-disclosure were operationalized by word count and depth of participant response (Collins and Miller, 1994). A one-way MANCOVA was conducted to test if word count, disclosure depth, user satisfaction, intention for future user, likeability, and perceived safety differed based on robot disclosure conditions, with the covariates of perceived stress, shyness, and NARS. Thematic and content analysis were used to explore the qualitative data (e.g. conversations with the robot and interview responses) (Richards, 2014).

**Results:** Interaction time ranged between 2 minutes 21 seconds and 9 minutes 54 seconds with an average of 4 minutes and 20 seconds. Manipulation check indicates that 56% of the participants failed to correctly identify the type of robot disclosure they experienced.

There was no statistically significant difference among robot disclosure conditions on the combined dependent variables after controlling for perceived stress, shyness, and NARS,  $F(12, 50) = 1.689$ ,  $p = .098$ , Wilks’  $\Lambda = .506$ , partial  $\eta^2 = .288$ . However, perceived stress significantly predicted word count,  $F(1, 30) = 5.674$ ,  $p = .024$ , disclosure depth,  $F(1, 30) = 4.839$ ,  $p = .036$ , user satisfaction,  $F(1, 30) = 8.910$ ,  $p = .006$ , likeability,  $F(1, 30) = 5.376$ ,  $p = .027$ , and future intention to use,  $F(1, 30) = 12.455$ ,  $p = .001$ . NARS also significantly predicted for intention for future use,  $F(1, 30) = 5.879$ ,  $p = .022$ .

Preliminary qualitative analysis of the post-interview found themes, such as: desire for deeper response, attribution of childlike quality, and comparison between robot and human companionship. Content analysis showed that 38% participants ( $N = 14$ ) expressed desire for deeper, more personal robot responses.

**Implication:** The current study provides several future research directions on social robot communication design and stress intervention using HRI. Firstly, the current results expand upon the findings by Martelaro et al. (2016) by testing different types of robot disclosures with an adult sample. The result of no significant difference in reciprocal disclosure and high error rate in the manipulation check for conditions suggest that adults and high schoolers might perceive and behave towards a social robot differently. Future studies will examine how age affect perception of robot disclosure. Secondly, the finding of perceived stress as significant predictor for multiple outcomes suggest that HRI design for stress interaction will need to take account of perceived stress level of the population, especially those with high stress. Lastly, our preliminary qualitative observations suggest that robot disclosures is not only important as an opener, but also as a response. Future studies on robot disclosure should investigate on

the effect of location of which disclosure occurs within a specific dialogue.

*Honson Ling (University of Washington)*

## Human-Technology / Brain-Machine / Human Systems Integration

### F-1

#### Exploring the Feasibility and Implications of Augmented Reality Brain-Machine Interfaces

Companies worldwide are investing in augmented and mixed reality technologies. While current consumers typically interact with augmented reality using smartphones, companies are actively working to develop consumer-adopted augmented reality headsets or smart glasses. Augmented reality headsets offer exciting new user experiences by allowing unencumbered hand movements to interact with real and virtual objects and have the potential to create more immersive experiences. However, augmented reality headsets also bring the potential for increased distraction, as virtual objects and information may automatically appear within a user’s field-of-view.

One major challenge in augmented reality is to determine when and where digital information should appear. These decisions are important for both user safety and cognitive ergonomics. From a safety standpoint, augmented displays must avoid obstructing important real-world objects, such as tripping hazards or oncoming cars. Augmented reality companies are actively developing systems that recognize and react to real-world objects. Regarding cognitive ergonomics, digital information must be strategically presented to avoid mentally overloading the user. Existing design principles for human-computer interaction may face limitations when users are wearing augmented reality headsets, which may create more immersive and distracting experiences compared to smartphones (while noting that smartphones are already extremely powerful at drawing users’ attention). In an ideal augmented or mixed reality scenario, virtual objects or information would naturally appear or expand when a user seeks to interact, and then naturally fade away when the user is not engaged. In order to create these natural interactions, the augmented reality system needs some knowledge of the user’s cognitive state (e.g., focus of attention, alertness, etc).

In this work, we discuss strategies to predict a user’s cognitive state based on existing bio-sensors, emphasizing eye-tracking and electroencephalogram (EEG) and present a prototype hardware configuration. First, we review the brain-machine interface literature as applicable for integration with augmented reality headsets, including the current scientific limitations of cognitive state detection using non-invasive technologies. Then, we discuss the specific challenges that are introduced or amplified by the use of augmented reality headsets in contrast to other brain-machine interfaces, including increased motion artifacts and electrical artifacts from display hardware. Third, we present a preliminary system that wirelessly integrates a Magic Leap ML1 augmented reality headset with an Enobio research-grade EEG (see Figure 1). Using this system, we characterize electrical artifacts and demonstrate a pipeline for interaction with simple virtual objects.

In addition to investigating the potential of augmented reality brain-machine interfaces, we raise important questions for

discussion about if and how researchers should proceed with the development of technologies that can detect a user's cognitive state. These questions include: 1) what bio-sensor data should be available for consumer products, 2) who should own these data, and 3) if we can accurately predict a user's cognitive state, are there limitations on how this information can ethically be used?

*Sam Michalka (Olin College of Engineering)*

## F-2

### The Role of Clinical Psychologists in Human Biotechnology Enhancement

**Statement of Problem:** This poster will focus on clinical psychologists' role in human biotechnology enhancement, a gap of research that currently exists in both the fields of psychology and technology. As biotechnology sciences advance into the future, people will have more opportunities to selectively enhance their genetics, body, brain, and mind. Many benefits, such as longer life-span, increased capabilities, and alterations of appearance are becoming possible (Miah, 2016). In contrast, enhancement technology may pose increased long-term health risks, negatively impact mental health, create unwanted side-effects, and narrow the life-span. Either way, this technology will undoubtedly impact our culture, society, the human condition, and how humans function on a daily basis (Hoffman, 2017). Clinical Psychologists can be greatly used to help individuals navigate these issues as humanity draws closer to a "trans-human" era. Historically, Clinical Psychologists have played a major role in rehabilitation of physical and psychological deficits. They have taken part in the care of patients going through transformation processes, such as plastic and reconstructive surgery, bariatric surgery, performance enhancing drugs, implantation of pain-relief devices, as well as those undergoing sex reassignment. However, it appears little has been discussed on how Psychologists may assist those electing to undergo human enhancement transformations. It is believed the same skillset Clinical Psychologists possess may be useful help these individuals navigate the transformation process and improve outcomes. This poster will review several methods in which Clinical Psychologists of the future can play a major role in the education, ethics, policy-making, research, evaluation, rehabilitation, and decision-making process of individuals who choose to undergo biotechnology enhancement.

**Procedures:** The poster will emphasize articles discussing the application of biotechnology for human enhancement and individuals electing to transform their genetics, body, brain, and/or mind. We will access major databases, including PsychINFO, Web of Science, Pubmed, and Google Scholar to acquire data on the topic. The poster will use a strategic scoping strategy developed by Arksey and O'Malley (2005) to examine the current body of literature in effort to examine existing research, illuminate gaps in published data, and generate new ideas for how Clinical Psychologists can play a role in human enhancement.

**Analyses:** This poster will summarize current and relevant research related to human biotechnology enhancement and discuss methods in which Clinical Psychologists can be assistive in the future. We will examine the extent, range, and nature of research and apply qualitative thematic analyses to collate and synthesize our findings. Key themes and points of discussion will be identified within the data to formulate ways psychologists can be assistive in the field of

biotechnology enhancement. This will allow us to better understand how individuals can utilize Clinical Psychology to improve psychological and physical outcomes.

**Results:** The poster will present results from the literature review and identify a number of topic areas in which Clinical Psychologist may become integrated into the field and science of human enhancement technology. We will include a discussion of the psychological changes individuals who undergo biotechnological enhancement procedures may encounter, such as, emotional distress, poor adjustment, problems with acceptance, and negative alterations in functioning (Brey, 2008). Furthermore, several strategies and methodologies will be proposed, which Clinical Psychologists may find useful in the future to assist individuals electing to receive human biotechnology enhancements.

**Practical Implications:** The results of this review will provide important considerations exploring how Clinical Psychologists can play a role in the advancement of "trans-human" technology and enhancement (Miah, 2016). The results will allow further discussion of the implications in clinical practice as it relates to expert opinions. Many questions are being raised in regard to who will be allowed to make these decisions and how these technologies should be used. These advancements will allow humans to make incredible alterations to their capabilities and appearance. However, with any major change to one's life, body, and mind, one must expect an adjustment of some kind. Here enters the Clinical Psychologist. It is expected that technology in this area will continue to grow, thus it is important to consider ways in which Clinical Psychologists will become utilized and involved so that we may be better prepared for the future.

**Conclusions:** Several uncertainties regarding how biotechnological enhancement will impact the individual, society, and human condition remain (Masci, 2016). This poster will promote discussion regarding the role of the Clinical Psychologist in the advancement of biotechnological enhancement. It will also illustrate ideas for future research and new directions as the field of psychology and technology become more strongly connected. It is important for Clinical Psychologist to continue expanding their roles into new areas of care. As we enter a new paradigm, where humans may choose to "upgrade" their own capabilities, Clinical Psychologists should take action by considering the roles they can play in this technological revolution.

*Alex Cook (University of Indianapolis)*

## Judgment and Decision-Making

### G-1

#### Development and Evaluation of a Mobile Health Clinical Decision Support Tool for Patient Migraine Management

**Background:** Migraine is a common disease affecting more than 1 in 10 Americans (Lipton et al 2007). Its symptoms are highly disruptive to suffers, making migraine the most disabling neurologic disease in the world (Leonardi & Raggi. 2013). Migraine is a chronic condition with episodic symptoms, which contributes to the complexity of patient decision-making regarding migraine management. Migraine management has two components: preventive and acute management. Preventive management includes routine behaviors that reduce the frequency of migraine attack onset. These include taking

a preventive medication routinely, stress management, and having a consistent healthy daily routine (including high quality sleep, consistent eating and drinking, and aerobic activity). Acute management requires in-the-moment decision-making to determine the correct differential diagnosis for any given headache attack, and then to enact the appropriate acute management plan; appropriate acute management can include a variety of acute medication components as well as behavioral strategies (Ramsey et al, 2018). Taken together, this series of decisions can be conceptualized as “adherence” to acute and preventive migraine management plans. Adherence to migraine management plans takes place outside of the doctor’s office. Personal data-based clinical decision support tools (CDSTs) represent a unique opportunity to improve adherence to patient migraine management as a just-in-time intervention.

**Methods:** The current study describes the development of a CDST for patient migraine management embedded in a HIPAA-compliant iOS application containing questions validated in prior research. Participants were recruited from a research registry of patients with a confirmed diagnosis of migraine. Each participant completed tasks in the app for a minimum of 7 days. Participants then participated in individual interviews with a health psychology and/or health education researcher, with structured questions around the following categories: user experience, content, burden, and overall satisfaction. Themes that emerged from the interviews are described below.

**Results:** 10 participants (M age = 44.2, SD = 15.7; 100% female; 60% White, Non-Hispanic, 20% White, Hispanic, 1% Black, Non-Hispanic, 1% Black, Hispanic; M Headache days/30 days = 10; SD = 3.8) evaluated the diary alone; 11 participants (M age = 37.7, SD = 7.9; 81.8% Female; 90% White, Non-Hispanic, 9% Asian; M Headache days/30 days = 12.6, SD 6.2) evaluated the diary and embedded CDST intervention. Key findings were:

**User Experience:** Patients preferred detailed instructions for using the app, and graphic representations of their data presented back to them weekly.

**Content:** Patient report suggested 1) information about taking medication early was particularly helpful; 2) focusing on lifestyle change was important; 3) bringing each piece of information back to migraine is essential for patient buy-in to the rationale for behavior change;

**Personalization:** Participants reported that 1) personalization of adherence information improved utility, 2) personal input in goal-setting improved buy-in, and 3) personalization of timing and method of data input improved diary adherence.

**Burden:** To reduce burden, patients provided feedback that the data collection should be: 1) adaptable, so that if their schedule changes they can change when they interact with the app, 2) brief, 3) similar each day so that they can memorize the questions, and 4) minimal in midday interaction requirements of the app, which was challenging for patients. Patients also reported alert fatigue with reminders, and suggested integrating preventive reminders into the data collection procedure itself; just-in-time reminder notifications would ideally only be used sparingly, and when the just-in-time nature is essential to their function.

**Satisfaction:** Most patients reported being satisfied with the app, in particular 1) they liked getting new information about their personal migraine patterns, 2) they reported that tracking behavioral management factors alone seemed to improve their adherence, and 3) they thought the information provided was important, foundational

information to which each person with migraine should be exposed early on in their treatment. Patients who were not satisfied with the app disliked it because of 1) burden, 2) reminder fatigue, and 3) prior exposure to the information.

**Conclusions:** Patients require support and guidance while initiating mobile health applications for research. Design features must be designed to support understandability and information processing. In this study, participants reported that both the CDST intervention, as well as merely self-monitoring, improved their migraine management behavior. These findings support the value of self-monitoring as a potential behavior change intervention, but also highlight the challenges to designing nonreactive outcomes assessment. These early findings suggest significant growth potential for digital migraine management support. This application also represents for patients an easily accessible electronic health record to facilitate long-term decision making around their migraine occurrence and care, parallel to services already provided for clinicians in a patient’s electronic health record.

*Ivy Gosnell (Ferkauf Graduate School of Psychology, Yeshiva)*

## G-2

### **Nature and Machine Heuristics: Activation and Dominance in Response to Organisms, Cyborgs, and Machines**

Heuristics—mental shortcuts that quickly but non-optimally facilitate decision-making—drive much of human behavior; although efficient, reliance on these shortcuts can lead to systematic logical errors or distortions in perception known as cognitive biases (Tversky & Kahneman, 1974). These can include errors in evaluating people or technological agents. One such class of bias relies on heuristics associated with a thing’s ontological category, or the kind of thing it is thought to be (see Kahn et al., 2011). Two ontological category biases emerge as particularly important for evaluations of machine agents. First, the machine heuristic is an assumption that machines are objective and so are more credible in information tasks (Sundar, 2008). Second, nature bias is a tendency to evaluate something natural as good and unnatural things as bad (see Allchin & Werth, 2017), reliant on a heuristic that elements of the natural world are morally, aesthetically, or socially normative (Daston, 2014) or nonspecifically and inherently better according to their purity (Haidt & Joseph, 2004). In sum, the former emphasizes superiority of the technological and the latter emphasizes superiority of the natural.

Although these mental shortcuts are conceptually divergent, it has not yet been examined whether they are empirically divergent with respect to agents of natural, machinic, or mixed-feature categories, and with respect to the technical or naturalistic behaviors. In short, our understanding of these efficient-but-flawed agent-perception processes may be marred by an entangling of agent and behavior perception. Following, we ask: RQ1: Does the agent’s cued ontological category influence the activation of nature bias and/or the machine heuristic? RQ2: Does the nature of the agent’s task influence the activation of nature bias and/or the machine heuristic?

**Method:** An online experiment is currently in pilot testing (with data collection and analysis completed by the conference) to address the posed questions regarding the ostensibly competing heuristics’ activation by agent category or task type. The study employs a 3x2 between-subjects design in which participants read a news story

about an innovation in which a particular agent (natural, cyborg, machine) solves a problem through a particular approach (information-driven or natural process-driven). Participants (to be recruited through campus convenience samples and social media outlets) will then be asked to make evaluations about the agent and about the problem solution quality and ethics, and then to complete demographics items and reflect on the story.

**Stimuli:** To examine the intersections of technological and naturalistic influences on triggered heuristics, stimuli relied on an adaptable narrative in which agents and tasks could meaningfully vary in those properties. Specifically, participants view a news article with accompanying photograph (validated by time spent on the displaying survey page and an attention check) focused based on an actual magazine article (Simon, 2018) about a “robotic pollinator” (i.e., a machinic agent completing a natural task). The basic content of the article was adjusted to (a) represent each agent/task combination and to (b) remove confounding language that could trigger the target heuristics inadvertently. The agents were a natural bee, a cyborg bee (Paez, 2018), and a robotic bee (a variation on the original story, for morphological parity across conditions). Stories featured an innovation in which the bees help solve the problem (declining crops in the face of population increases) through natural solutions (innovative pollination) or information-based solutions (information gathering toward farmer adaptivity) according to the agent’s abilities.

**Measures:** Participants will complete measures representing evaluations of superiority across both naturalistic and technical tasks (i.e., sentiments signaling that leaning toward naturalistic or machinic bias activation). They were: the competence dimension (capable, responsive, interactive, reliable, competent, knowledgeable) of the Robotic Social Attributes Scale (RoSAS; Wyman & Stroessner, 2017), and a general evaluation of the solution to the pollination/information problem (13 items reflecting effectiveness and elegance; adapted from the Creative Solution Diagnosis Scale; Cropley & Kaufman, 2012). Additionally, per Bellur and Sundar (2014), activation of machine, nature, and authority heuristics (each by seven 7-point semantic differentials) were measured. Participants also responded to control items: (a) biospheric, egoistic, and altruistic values (i.e., pre-existing pro-nature, -self, or -society attitudes; 13 items adapted from Steg et al., 2012), (b) a single-item measure of the perceived ethicalness of the bee solution, and (c) demographics. Additionally, an attention check was employed (indicating the problem being addressed in the story) along with a manipulation check (single semantic differentials rating the agent as mechanical to natural and the solution from natural to technical).

**Plan for Analysis:** Evaluations of the agent and the solution will be compared across conditions (controlling for subject values and ethical evaluation) to determine the conditions under which each mental shortcut is activated, and to explore the potential for one to eclipse the other. Open-ended responses for thoughts on the news story will be inductively explored to contextualize findings. Data collection and analysis will be completed well in advance of the conference.

*Jaime Banks (Texas Tech University)*

### G-3

#### Understanding Digital Nudges from an Analytical-Behavioral Perspective

With the recent development of smartphones and mobile devices, the context in which humans behave has been changed. Many of our

decisions are now taken in a digital framework, and everyday we take an increasing number of digital decisions. As we do not always take the optimal decision, the one that maximizes our gains, and cannot afford to take time to analyse all possibilities at their complete extent, we resort to heuristics and biases to judge alternatives. In a digital environment, to favor such decisions in their intended purpose, UX and UI designers implement multiple interventions in webpages and mobile apps to help users and consumers reach the final target response, according to Human-Computer Interaction (HCI) techniques and principles. In order to facilitate user’s response, designers use what has been recently called as digital nudges (Weinmann, Schneider & vom Brocke, 2016; Schneider, Weinmann & vom Brocke, 2017). Nudges are classically defined as the context which alters one’s behavior in a particular way (e.g. Thaler & Sunstein; 2009). Although nudges still remain to be completely effective in real life contexts (Hummel, Toreini & Maedche, 2018), literature on digital nudges is growing, in part because it is easier and cheaper to implement digital nudges as compared to offline ones. Most of those studies use contributions taken from cognitive psychology, such as heuristics and biases. Behavior analysis, however, with over a century of valid scientific research, has gone unnoticed and behavioral explanations have not been offered enough as an alternative to understand how people behave in online contexts (Simon & Taglibue, 2018). On behavioral perspective, heuristics and biases are not considered to cause or to determine behavior, as any other inner cause (Skinner, 1953). According to behavioral analysis, they can be considered simply as descriptions of how people tend to make decisions (e.g. Simon & Taglibue, 2018). Causes are provided ultimately by the interaction between the subject and the environment one belongs (Skinner, 1953; Todorov, 2007). The behavioral perspective model (BPM) has been conceived as an approach to understand, interpret, and predict consumers behavior. It is based on behavioral principles, as an option to the use of cognitive processes embedded in marketing studies (e.g. Foxall, 2001; 2010; Oliveira-Castro & Foxall, 2005). One of the model most important contributions is to include positive and negative consequences simultaneously in Skinner’s triple contingency (e.g. 1953). The objective of this poster is to offer a behavioral perspective explanation to some online heuristics and biases, used as nudges to facilitate user’s online choices. Using BPM and behavior analysis principles (e.g. Baum, 2004; 2010), six cognitive effects will be explored and interpreted. Mirsch, Lehrer and Jung’s (2017) systematic literature review identified nineteen most frequent psychological effects cited in literature. Only the top six of them in terms of frequency in paper reports (framing, status quo bias or default choice, social norms, loss aversion, anchoring and adjustment and hyperbolic discounting) will be explored in this poster, due to length restrictions. The study will show how these effects refer to different parts of user’s behavior contingency. Some affect the antecedent aspects (divided into elements of individual learning history and online settings as proposed by BPM); other effects focus on responses, affecting behavior itself; and some effects may alter the power of consequences, either reinforcement or punishment, and either utilitarian or informational consequences (e.g. Foxall, 2010). The digital context provides a contingency, that is a space in which behaviour, subject, and possible positive or negative consequences of the response interact in a functional relationship. In this perspective, nudges can be understood as motivating operations (e.g. Fagerström, Foxall & Arntzen, 2010; Michael, 1982; 2000), or events that attribute value to one course of action or to the consequences of those actions, therefore increasing or reducing the probability of such

response to happen. To do so, these six effects will be transferred to the digital environment, and online examples will be shown to illustrate the theory. This is a preliminary investigation towards a more comprehensive HCI research using behavioral analysis principles. The study is underway on its analysis and on gathering digital examples. The paper may offer a contribution to rethink procedures and to design more effective and friendlier human-computer interfaces.

*Patricia Luque Carreiro (Patricia Luque Psychology)*

#### G-4

##### **'It's Written All over Your Face'—Applying Deep Neural Networks to Explain Human Mating Based on Facial Characteristics**

Human mating is subject to research in economics (matching theory), psychology (matching hypothesis), and computer science (matching algorithms). First, a rich body of research in economics examines human mating in online dating markets. Thereby, human mating is primarily explained by text-based (socio-) demographic characteristics and personality traits (Hitsch, Hortaçsu, and Arieli 2010). Facial characteristics, however, are not considered in those investigations. Second, a rich body of research in psychology investigates human mating (for a review, see: Finkel et al. 2012). Similar to the aforementioned contributions in economics, research in psychology majorly referred to text-based (socio-) demographic characteristics and personality traits to explain human mating in online dating markets. However, one notable exception should be stressed (even the investigation referred to offline dating). In particular, Valentine et al. (2014) analyzed human mating habits based on facial width-to-height ratios (FWTHRs) in a series of seven speed dating events (78 men, 81 women) concluding that “[...] higher [FWTHR; adjusted] are attractive to women for short-term relationships” (p. 806). However, Valentine et al. (2014) disregarded significant facial characteristics (e.g., averageness) that were defined as decisive for human visual perception (Rhodes 2006). Finally, human mating is subject to research in computer science on matching algorithms. In particular, companies such as Google and Microsoft patented technical procedures that target to match humans in online dating markets based on similarities of facial characteristics (e.g., US8670597B2 or US20130156274A1). However, besides detailed descriptions of technical procedures for the computation of facial characteristics, those patents do not provide information on the underlying rationale for particular matching algorithms and, most important, miss sound evidence on the relation of facial characteristics and human mating. Referring to the aforementioned shortcomings, this research analyzes human mating based on facial characteristics.

The analysis references on data of approx. 640,000 users of an online dating platform (period of data collection: 01/2018-07/2018). For the set of users, facial images (I), profile section data (II), and transaction records (III) are accessible. (I) The facial images are (pre-) processed with Convolutional Experts Constrained Local Models (Baltrušaitis et al. 2018) to extract, inter alia, facial landmarks (2D / 3D) and rigid- / non-rigid shape parameters. This research references on the most common measure to typify facial characteristics: Facial Width-to-Height Ratios (FWTHRs) following the approach suggested by Stirrat and Perrett (2010) (ratio of the geometric distance between the facial left and right zygion (facial width) and the geometric distance between the facial upper lip and upper eye (facial height)). Additional facial measures (as well as

pixel-based image information) will be added in a later stage of the research project. (II) The profile section data include information on users' physical characteristics such as body type and body height as well as demographic characteristics such as annual income, education level, number of children, and age. (III) The transaction records include information on the type and time of transaction.

(Simple) GLMs are applied to examine human mating based on facial characteristics. The dependent variable is defined by the number of contact requests for a given user (whereby only the initial contact request within a given dyad for a given pair of two users is considered; controlled for other user activities). The set of independent variables refers to the previously outlined facial, physical, and demographic characteristics, whereby various model-specifications are estimated. Further, a reference to samples that include all, female and male, users, only female users, and only male users as well as various sample modifications with regard to rigid- / non-rigid shape parameters are included.

The results indicate that facial characteristics enfold significant effects on human mating. In particular, an increase of the FTWHR leads to a significant decrease of contact requests for all, female and male, users ( $p < .001$ ). This significant negative effect of the FWTHR on contact requests can also be observed for only female recipients ( $p < .001$ ) (i.e., women are more often matching target if the FTWHR decreases). Contrarily, the FWTHR has a significant positive effect on contact requests for only male recipients ( $p < .001$ ) (i.e., men are more often matching target if the FTWHR increases). This finding holds when controlling for physical and demographic characteristics as well as for various alterations of rigid- / non-rigid shape parameters.

This research advances most recent research on human mating in economics, psychology, and computer science. Whereas previous research neglects facial characteristics (economics), disregards relevant facial characteristics (psychology), or misses evidence for the relation of facial characteristics and human mating (computer science), this research provides a comprehensive perspective on human mating based on facial characteristics. Correspondingly, this research assists market designers and software developers in improving existing approaches and implementation provisions for human matching in online dating markets.

*Daniel Boller (University of St.Gallen)*

## Learning / Education

#### H-1

##### **Using Machine Learning to Predict Activity Types and Social Interaction Levels of Children with High-Functioning Autism in Virtual-Reality-based Social-Skills Training**

Introduction: Autism is a longitudinal mental disorder that 1 in 59 children are diagnosed in the U.S. Especially, children with high-functioning autism (HFA) demonstrate average cognitive abilities but face social-interaction impairments, including interpersonal communication and social manner deficits. Research has introduced ways to enhance HFA children's social skills through deliberate training interventions. Naturalistic intervention has been one of the major training designs that encourage HFA children to practice their social interactions in experiencing either daily-routine or personally-preferred activities. Naturalistic intervention focuses on promoting

students' social engagement via diverse routine-like social contexts. In particular, naturalistic intervention has recently adopted virtual reality (VR) that aims at offering variant and diversified social simulations that feasibly occur in the real world (Didehbandi, Allen, Kandalaf, Krawczyk, & Chapman, 2016). Evidence supports that VR-based social-skills training enhanced HFA children's social skills (Ke & Im, 2013). Due to HFA children's dynamic psycho-social profiles, it is difficult to identify optimized intervention activities or features to stimulate students' social skills in the interventions. Using machine learning, this study aimed at identifying proper activity types for HFA children and estimating their social-interaction levels in VR-based social-skills training. The study addresses a key research question: How do variant machine learning methods predict proper activity types for HFA children and their social-behavior level during VR-based social-skills training?

**Method: Samples:** This study sampled a total of 51 one-hour video-archived sessions for two children with HFA from VR-based social-skills training. Each participant attended multiple one-hour sessions over 8-16 weeks.

**VR-based social-skills training.** Using Opensimulator, we designed and conducted VR-based social-skills training. Opensimulator is an open-ended VR platform that supports multiple users' social interactions that encourage learners to experience social simulations. The participants of this study were asked to perform social acts tailored to game-like activities under variant social scenarios. There were several social scenarios of the activities: (1) house and building constructions considering a client's needs, (2) designing a building in a novel way, (3) playing a math-puzzle game, and (4) group discussion in a storytelling activity. Two or three facilitators also joined and provided in-time scaffolds (e.g., resolving technical issues in VR and giving backdrop narratives to the participants) to support students' social acts during the training. This training intervention aimed at five social skills: (a) responding and (b) initiating social interactions, (c) interpersonal negotiation, (d) self-identity, and (e) cognitive flexibility.

**Data processing and analysis:** Before employing predictive machine learning of this study, systematic behavior observations were conducted to code the participants' social interaction behaviors by using time-sampling (per 30 seconds). To keep high inter-reliability of the coding results, three experienced coders initially sampled 20% of training session data with inter-rater reliability higher than .85, and then refined their coding results until their agreement reached 100% among themselves. Table 1 is the dataset structure for classification algorithm implementations (refers to the appendix). Furthermore, we coded the participants' activity types, as well as the occurrences of students' social skill performance to build machine learning classifiers. To run the data analyses, we used four baseline classifier algorithms (i.e., Decision Tree, Random Forest, Neural Network, and Naïve Bayes). We implemented a total of 10 stratified cross-validations to ensure that the sample of the data is a good representative of the whole data population.

**Results:** In regard to detecting the activity types for HFA children in the social-skills training, the study result confirmed that the classifier Decision Tree performed best with the accuracy of 39.9%. All the statistics results from the classifier implementations are shown in Table 2 (refers to the appendix). With regard to predicting participants' social-interaction performance level, the classification accuracy of Naïve Bayes and Random Forest were higher than those of other classifiers. The results of the classifier implementations

confirmed that Naïve Bayes better performed on the initiation of social interactions (61.4%), interpersonal negotiation (82.9%), and cognitive flexibility (69.4%). Also, it was noted that the classifiers Random Forest and Decision Tree attained the higher accuracies on responding (96.4%) and self-identity (86.1%), respectively.

**Discussion and Implication:** This study used machine-learning classifiers to detect proper activity types and social-behavior level of HFA children in VR-based social-skills training. Contrary to the low prediction result as to the activity types, the classifiers appeared highly accurate to predict students' social interaction performance. In future studies, to increase the accuracy of the classifiers on proper activity types of VR-based social-skills training, we will include additional log-initiated metrics that are related to students' social skill implementations. Also, this study will specify the level of classifier variables to detect proper intervention settings of VR-based social-skills training more accurately.

*Jewoong Moon (Florida State University)*

## H-2

### Exploring Collaborative Gameplay in Virtual Reality-based Social-Skills Training for Children with Autism

**Introduction:** Research on embodied social cognition reports that individuals with autism have difficulty understanding various social clues in social circumstances. This socially-disembodied disposition of children with autism is likely to lower their social interaction skills (Eigsti, 2013). In response, social-skills training in virtual reality (VR) is a purposeful intervention to strategically improve social competencies of children with autism in a 3D-simulated social world. VR in social-skills training allows children with autism to experience different types of social contexts to practice their social interactions. However, questions remain about what design and implementation strategies of the VR-based learning environment effectively promote social skills of children with autism. Prior research has introduced collaborative gameplay as an interactive learning strategy that motivates children with autism (Sung & Hwang, 2013). Empirical research on the impact of collaborative gameplay in developing social competencies of children with autism, however, is still limited. This study aims to examine the impact of collaborative gameplay on social skills development of children with autism in a VR-based learning environment. The research question of this study is: Will collaborative gameplay better promote social skills of children with high-functioning autism than individualistic gameplay?

**Method: Samples and intervention design:** This study collected a total of 42 one-hour videos (Collaborative = 22, Non-collaborative = 20) of social-skills-training sessions hosted for two 10-14-year-old children with high-functioning autism (HFA). All training sessions in VR were screen-recorded. Each participant attended the social skills training over 16-26 intervention sessions. The number of the training sessions varied based on each participant's learning progress.

Using Opensimulator, we built a 3D virtual environment as a simulated social world that offers multiple "sandbox" social scenarios for children with autism. The VR-based social learning environment enables children with autism to experience various types of collaborative gameplay activities, including (a) design quests, (b) scavenger hunting, (c) storytelling, (d) roleplaying, (e) social problem-solving, and (f) math games. Each target child participated in these gameplay activities along with two to three peers in the virtual world. Multiple

non-player characters were puppeteered to provide naturalistic, environmental prompts to the participants. The prompts were to stimulate the participants' practice and portrayal of the targeted social skills during collaborative gameplay. The intervention targeted five core social skills: (1) responding, (2) initiating social interactions, (3) interpersonal negotiation, (4) self-identity, and (5) cognitive flexibility.

**Data collection and analysis:** We conducted a systematic behavior analysis of the participants' video-recorded training sessions. Based on the ethogram of the behavior analyses, we coded the participants' social-interaction behaviors using time sampling (per 30 seconds) as the major unit of analysis. Supported by a self-developed behavior observation tool, we coded the occurrences of each type of the aforementioned social-interaction behaviors in both collaborative and non-collaborative gameplay sessions. Three trained coders independently coded randomly-selected training sessions and then compared the results of the behavior coding. The inter-rater reliability of the coding was higher than .80. The coding was then iteratively refined until all the coders reached 100% agreement among themselves. This study implemented one-tailed t-test to identify the effect of collaborative versus competitive gameplay on the learning outcomes (i.e., the average occurrences of social-interaction behaviors over time).

**Results:** According to the one-tailed t-test, there is a statistically significant difference in the performance of cognitive flexibility between the collaborative gameplay ( $M = .44$ ,  $SD = .24$ ) and the individualistic gameplay ( $M = .30$ ,  $SD = .26$ ) conditions,  $t(40) = 1.79$ ,  $p = .04$ . There were no significant differences on other social skills as Table 1 shows (refers to the appendix). Figure 1 demonstrates a histogram of social skills improvement of the HFA participants and a heatmap portraying cognitive flexibility improvements over time (refers to the appendix). The red-color cells in the heatmap represents the higher performance of cognitive flexibility whereas those of blue-color indicates the lower cognitive flexibility performance. The heatmap shows that there are more red-color cells in the collaborative gameplay condition than those in the individualistic gameplay, and HFA participants show gradual improvement in cognitive flexibility performance in the collaborative gameplay condition.

**Conclusion and implications:** The study finding demonstrates that collaborative gameplay in VR-based social skills training better promoted HFA participants' cognitive flexibility. This exploratory study finding implies that collaborative gameplay helps to enhance HFA participants' dynamic perspective-taking and adaptive thinking in the given social and design problem-solving circumstances. Further research is warranted to further investigate how collaboration and peer interactions of children with autism occur and promote cognitive flexibility across various collaborative gameplay tasks.

*Jewoong Moon (Florida State University)*

### H-3

#### Exploring Students' Peer Interactions and Gameplay Patterns via Sequential Analysis

**Introduction:** Research on game-based learning (GBL) has explored vital design factors to promote students' meaningful gameplay (Abdul Jabbar & Felicia, 2015). Evidence supports that students' meaningful in-game actions represent their deeper understandings of game tasks relating to learning (Holbert et al., 2014). Hence, to

understand students' learning trajectories in a game, it is essential to identify the pattern of students' gameplay sequences that emerge.

Social constructivists have stated that students' peer interactions are possibly related to students' meaningful play because students' collaborative knowledge-sharing practices in GBL can enhance their learning (Corredor, 2018). Although some GBL studies have favored collaborative environments (Ke, 2008), in-depth explorations are still limited regarding how students' outside-game peer interactions are related to their meaningful play in a learning game. In this study, we implemented sequential analysis to explore students' gameplay sequences across peer interaction levels. The research question of this study is: How do students' gameplay sequence patterns differ by students' peer interaction levels (high- vs. low-interaction) when playing a math game?

**Method: Samples and game:** This study sampled a total of 102 gameplay sessions played by 32 middle school students from a public school in North Florida. All the gameplay sessions were video- and screen-captured. We implemented E-Rebuild, a single-player 3D architecture simulation game that encourages students to rebuild buildings and structures in disaster-occurred sites (Ke, Shute, Clark, & Erlebacher, 2019). Students are requested to investigate the damaged sites and fulfill house or shelter constructions under a limited budget. This game budget allows the students to purchase their building materials (e.g., shipping containers, bricks, flooring, and a door). This game is composed of multiple game episodes: (a) island, (b) desert, and (c) urban school. Aligned with Common Core State Standards for middle-school mathematics (Porter, McMaken, Hwang, & Yang, 2011), students in this game are asked to understand the ratio and angle measures via completing key embedded design quests: (1) trading building materials, (2) constructing shelters, and (3) allocating refugees to given shelters.

**Data processing:** Using the analytic tool BORIS (Friard & Gamba, 2015), we conducted systematic behavior observations to estimate student's time duration of peer interaction. Through open-ended coding, an initial behavior coding scheme was developed by multiple expert E-Rebuild players who were also knowledgeable in GBL research. Table 1 is the behavior coding scheme of the analysis (refers to the appendix). After attaining a successful inter-reliability of the coding results, the coders then iteratively compared their coding results until they meet 100% agreement of their coding. We then calculated gameplay sessions' peer interaction efficiency (PIE) scores based on the formula shown in Figure 1 (refers to the appendix).

**Data analysis:** Using sequential analysis (Bakeman & Quera, 2011), this study extracted frequent gameplay sequences with the likelihood higher than the preset threshold value ( $Pr > .80$ ). This study clustered two groups by resampling gameplay session data depending on PIE scores (i.e., the upper and lower 20% of gameplay sessions) and then compared gameplay sequence patterns across the two peer interaction levels (high- vs. low-interaction).

**Results:** The sequential analysis captured a total of 15 frequent gameplay sequences performed by the participants (high interaction = 7 / low-interaction = 8) as shown in Table 2. According to the results, there were noticeable differences in students' gameplay sequence patterns. First, the low-interaction condition had more chances to experience learning engagement (i.e., content engagement) in their gameplay than those of the high-interaction one. Second, the low-interaction condition shows that the students tended to perform help-seeking actions during their gameplay (e.g., Using embedded learning support or asking help from facilitators)

compared to those from the high interaction one. Lastly, in the low-interaction condition, students were likely to succeed in their game tasks compared to those of the high-interaction condition. In other words, the students' in-game help-seeking actions enabled the students themselves to revise their initial solutions to the game tasks (e.g., the 3rd and 4th gameplay sequence patterns from the low-interaction condition from the appendix table). The iterative revisions of their game solutions might lead to their success in the game task performance.

**Conclusion and implications:** This study revealed how students' gameplay sequence patterns differed by peer interaction levels. The study result confirmed that the low peer interaction outside the game was associated with a higher chance of learning engagement and game-task accomplishment. Via further analyses, we will examine how and why students' non-peer-interactions outside the game led to students' more meaningful gameplay related to learning. The study finding also suggested that educational game designers should understand how specifically students' peer interactions influence meaningful play in GBL.

*Jewoong Moon (Florida State University)*

#### H-4

##### **Better Science Through Diversity: When Social Scientists, Geoscientists, and Silicon Valley Collaborate to Create Diverse, Equitable, and Inclusive University Geoscience Departments**

Thriving workplaces are ones in which diverse groups collaborate to address complex, vexing problems that no single individual can tackle. In fact, over a decade of research has amassed to show that diverse groups are more innovative and productive than homogenous ones (Freeman & Huang, 2014; Page, 2007). Acknowledging this reality, leaders in industry and, to a lesser extent, higher education, have devoted countless resources to addressing diversity, equity, and inclusion (DEI; Dobbin & Kalev, 2016; 2018).

Yet, despite efforts to increase diversity in geosciences and across STEM fields, only 35 out of a total of 737 doctoral degrees in earth, atmospheric, and ocean sciences were conferred to underrepresented minorities in the year 2012 (National Science Foundation, 2015). This stark contrast has created a "chilly" or hostile climate, which stifles or damages collaboration and, ultimately, outcomes.

Does diversity training reduce prejudicial behaviors? Unfortunately, the weight of the evidence shows that they are largely ineffective at changing attitudes and behaviors (Dobbin & Kalev, 2018; Moss-Racusin et al., 2014). One possible explanation for this is that there is low engagement in these training programs, which typically deliver content or one-off workshops, and are compliance-focused and involve passive learning. Even recent pushes to use technology-based DEI training merely digitize the problems that are inherent in traditional DEI training, and therefore are unlikely to push the needle regarding DEI issues.

Given these challenges, for our proposed Applications Poster Presentation, we will outline how we integrated the two theoretical frameworks that guided our National Science Foundation-funded project with the innovative technological architecture behind our mixed-reality simulations. These simulations provide our participants with novel ways to learn and practice how to advocate for a more diverse, equitable, and inclusive workplace. Guided by social closure theory (Vallas & Cummins, 2014), and social cognitive theory

(Bandura, 1986), our project goes beyond merely delivering content regarding DEI and instead engages faculty members in authentic scenarios that can convert knowledge into actions that open up opportunities to those who have been underrepresented in geosciences. Our application of social closure theory targets gatekeeping decisions that close off opportunities, thereby changing policies and cultural practices that reward inclusive and equitable behaviors.

However, because mastering social dynamics that involve DEI issues is so difficult to do naturally in person (i.e., the situations do not present themselves naturally very easily, and the risks of failure while intervening can be prohibitive), we created a novel approach that combines the best of human intelligence/intuition with artificial intelligence (AI) to tackle this problem. The human-in-the-loop architecture that we developed creates for our learners the sense that they are experiencing something they are likely to face in their daily work lives, while the avatars create a psychologically safe environment to practice. To realize this architecture, a single professionally trained actor can "inhabit" up to five different avatars at once. Learners interact with the avatars based on the direction the conversation takes them. The avatars communicate with learners in real time and behave differently based on whether a learner is exhibiting a predetermined set of "strong" or "weak" behaviors. Although learners are interacting with real people, the learners are face-to-face with avatars, which allows learners to take social risks while practicing skills for creating better workplaces. AI capabilities built into the technology allow the actor to switch between characters and create nonverbal and verbal communication actions that provide learners with the sense that they are actually engaging with real people in an authentic social situation. In our poster presentation, we will go into much more depth about the technological architecture, and how we have seamlessly woven together human conversational and reasoning abilities with AI to create simulations that are both "intelligent" and engaging, as well as scalable for use in many different contexts.

For the intervention that we will report on, we recruited 30 tenured geoscience professors who demonstrated that they had significant "gatekeeping" responsibilities. These individuals participated in an intensive 3-day workshop where they learned skills to address prejudice, and completed the first of three simulations. Over the next academic year, these geoscientists participated in three online journal clubs to further their learning, and also in two more simulations. To demonstrate the impact of our intervention we will show results from the four waves of survey data regarding our participants' self- and collective efficacy. We will detail our Bayesian analytical approach to show the longitudinal changes over a 1-year period. We also recorded all participants' performances on the simulations. Although we are currently in the beginning stages of analyzing these audio/video data we will present our preliminary findings regarding the variety of approaches our participants took to navigate the social and political dynamics in our three simulations, and the qualitative changes our participants evinced after participating in these three simulations.

*Jason Chen (College of William & Mary)*

#### H-5

##### **Intelligent Tutoring Systems for the Filipino Learners: Current Research, Gaps and Opportunities**

For the last 50 years, intelligent tutoring systems had been a valuable educational resource for teachers, educational institutions, and

students (Kulik & Fletcher, 2016). With intelligent tutoring systems (ITS), students' specific learning needs are addressed appropriately which may not be possible for courses with large class size. Students are given equal opportunities to learn with the assumptions that these educational technologies are readily accessible. Students received individualized feedback (Schroeder, Adesope & Gilbert, 2013) and real-time assessment which does not require the presence of a human teacher (Bringula, Basa, Dela Cruz, & Rodrigo, 2016; Green, 2011; Marion & Oluwafunmilayo, 2011). ITS supports self-regulated learners where students can create their own learning space and time (Winne et al., 2006; Duffy & Azevedo, 2015). It provides a non-embarrassing learning environment since the interaction happens only between the ITS and the learners (Bringula, Alvarez, Evangelista, & So, 2017). It is now incorporated in mobile technologies allowing learning to be more accessible and flexible (Aker, Ksoll, & Lybbert, 2012).

However, most ITS (e.g., Matsuda et al., 2013; Graesser, Chipman, Haynes, & Olney, 2005; Kizilkaya & Askar, 2008) are mainly conducted in the context of foreign settings which are culturally different from the Philippines. It has been established that there is a link between emotions and cultural orientations of a person (Scollon et al., 2004; Kim-Prieto, Fujita, & Diener, 2004; Shaver & Schwartz, 1992). Furthermore, cultural differences are attributed to test anxiety (Cassady, Mohammed, & Mathieu, 2004) and reward allocation (Fischer & Smith, 2003). For instance, a Filipino student may not feel comfortable when a pedagogical agent is showing disgust or anger when a student committed repeated mistakes (Bringula, Fosgate, Garcia, & Yorobe, 2018) since Filipinos are known to be "mahiyain" (easily embarrassed) (Lasquety-Reyes, 2016) and "balat-sibuyas" (too sensitive) (Selmer & De Leon, 2001). Moreover, for a specific subject such as mathematics, the native language of the learners is taken into considerations (Abedi & Lord, 2001). Thus, a culturally-aware ITS is important since it understands the learners, it improves the quality of human-computer interaction (Blanchard & Frasson, 2005), and finally, it may lead to better academic performance.

While there are attempts to develop an ITS in a local setting (e.g., Bringula et al., 2017, 2018; Lagud & Rodrigo, 2010), very little is known on the cultural basis of ITS development. In light of these research gaps, this paper intends to investigate the state of ITS and mathematics learning studies conducted in the Philippines. Specifically, it aims to 1) present the state of ITS research in the Philippines, 2) present studies of mathematics teaching in the Philippines, and 3) propose an unexplored research topic. In this study, culture is delimited to the Filipino language.

An exhaustive literature search relating to ITS studies conducted in the Philippine setting will be conducted to answer the first objective of the study. Published papers in ITS in online databases (e.g., EbscoHost, ScienceDirect, Scopus, SpringerLink, Sage Journals, Taylor & Francis, Emerald Insight, ACM Digital Library, IEEE Digital Library, and Google Scholar) for the last 10 years (2008-2017) will be considered. Only journal articles and conference proceedings will be considered. Furthermore, journal publications (e.g., <https://e-journals.ph>) and conference proceedings of recognized computer societies (e.g., Philippine Society of Information Technology Educators [PSITE] and Computing Society of the Philippines [CSP]) in the Philippines will be included. The state of ITS research will be characterized in terms of number of published articles, researchers' profile (name, institution, number of published articles, name of journal publication,

Filipino collaborators, foreign collaborators, and affiliations), types of ITS (e.g., pedagogical agents, intelligent games, intelligent computer-aided instructions, etc.), strategies (e.g., textual feedback, facial feedback, synthetic emotions, and affect), and applications in subjects areas (e.g., mathematics, science, language, history, etc.). The phrases "intelligent tutoring systems", "pedagogical agents", "intelligent games", "intelligent computer-aided instructions" will be combined with keywords and phrases "Filipino", "Filipino learners", "Filipino students", "Philippines", and "Manila" will be used to search for articles. The same procedures will be conducted to answer the second objective with slight modifications with the keywords used (e.g., "mathematics" + "Filipino").

Our initial search results yielded to 39 ITS papers whose primary content is in Algebra (at least 20 papers) with the use of a pedagogical agent. There are 7 papers in mathematics learning focused on word problems and arithmetic taught in mother tongue. These findings suggest the wide gap in understanding the learners and their teachers in mathematics learning. Other foundational fields of mathematics, such as fractions, are not well-investigated. Therefore, there is a need for further studies that would further enrich the literature.

*Rex Bringula (Ateneo de Manila University)*

## H-6

### Females High School Students' Perceptions of the Value of Digital Math Games for Classroom Learning

Digital educational games have been increasingly incorporated into secondary classroom curricula (Papastergiou, 2009; Bourgonjon, Valcke, Soetaert, & Schellens, 2010; Jackson & McNamara, 2013). In fact, findings have shown that classroom game use enhances students' cognitive investment in a given academic subject and participation in classroom activities (Annetta et al., 2008 Iten & Petko, 2016). However, examinations of students' perceptions of game utility in learning and actual game use in classrooms is limited, particularly in STEM-based classes in which games are often used (Annetta et al., 2009; Papastergiou, 2008; Hamari et al., 2016). Studies of these perceptions are particularly lacking among female students who are more likely than males to show low performance in and negative attitudes towards mathematics (Dang & Nylund-Gibson, 2017). This study was designed to bridge this gap by investigating students' perceptions of actual digital game use in the math classroom and its perceived utility, and challenge among female students in single sex high schools.

Methods: Participants included 99 ninth-12th graders, over 45% of whom were females of color. All had been exposed to digital educational games in their math classes. All students were administered a 43 item survey, via Survey Monkey, that elicited demographic information about the students including their media use habits and digital game genre preferences, and questions that we developed addressing students' perceived utility of games used in their math class, the challenge that these games provided, and their teachers' use of games in their class. These questions were derived from the limited literature pertaining to students' perceptions of game play in the classroom (see Bourgonjon et al., 2010; Hamari et al., 2016; Mongillo, 2008) and were answered using a 5 point Likert scale (1 = Strongly agree; 5 = Strongly disagree).

Results and conclusions: An exploratory factor analysis yielded three factors with an eigenvalue greater than 1 that were distinguished

by strong factor loadings: usefulness of game play in the classroom (10 items including “I view educational games as a learning tool”;  $\alpha = .94$ ); perceived challenge of game play in the classroom (4 items including “I feel frustrated when I play educational games in the classroom”;  $\alpha = .73$ ); and teacher use of game play in the classroom (4 items including “Most of my teachers incorporate educational games in a lesson to help me learn material”;  $\alpha = .67$ ). Usefulness emerged as the strongest factor of the three. Correlational analyses including the three constructs showed significant relationships between usefulness and challenge ( $r = -.34, p < .05$ ) and between usefulness and teacher use ( $r = .29, p < .05$ ). These three factors collectively accounted for 51% of the variance (Usefulness = 33%; Challenge = 10%; and Teacher Use = 8%). Relationships between the demographic variables such as students’ current use of educational games, independent of classroom play, and usefulness were negatively correlated ( $r = -.30, p < .05$ ).

To examine potential grade differences between student responses on the scales, a MANOVA was performed with grade as the independent variable and the three factors as dependent variables. A significant effect of grade was found,  $F(3,77) = 4.42, p < .05$ , for evaluations of usefulness. Post hoc analyses showed a significant difference between 10th and 12 graders such that the former showed greater perceptions of game utility than the latter.

Overall, this study was designed to examine female students’ perceptions of educational game use in their classroom using a newly developed survey tool. The findings were consistent with those found in the literature as the strongest factors to emerge were those pertaining to the utility of game use, challenge of game play, and teachers’ game use (Bourgonjon et al., 2010; Hamari et al., 2016; Mongillo, 2008). Not surprisingly, students’ perceptions of usefulness of game play in the classroom were linked to their perceptions concerning their teachers’ use of games. However, unlike findings in the literature (Iten & Petko, 2018), perceptions of game challenge in the classroom were inversely related with game utility. Thus, games presenting greater levels of challenge may have been seen as less useful among these students. Collectively, these findings provide preliminary information concerning female students’ perception of game use in the math classroom which is likely to increase over time as games are increasingly incorporated into pre-college level curricula (Takeuchi & Vaala, 2014). Further replication and development of our survey with a larger sample and inclusive of greater numbers of items and constructs (e.g. math efficacy) also are clearly warranted. That students were not likely to see their own educational game use, outside the classroom as useful, raises interesting questions that we also plan to investigate. These goals are currently informing our follow-up work now underway.

*Karina Vargas (Notre Dame School)*

## H-7

### Helping Women Find Reasons to Stay in Computing Through Identifying Sources of Competence and Worthiness that Enhance a Healthy Self-Esteem Within Their Computing Major

Women are underrepresented in university computing majors, earning less than 20% of the bachelor’s degrees conferred in 2018. This lack of diversity has negative implications for future workforce teams and technology designs. While K-12 initiatives help encourage girls to try a major in computing, once enrolled in a computing major

university retention efforts are needed to help these women graduate and enter the technology workforce. Recognizing the connection between healthy self-esteem and academic achievement, previous research by this author used a two-factor model of self-esteem to understand how women’s self-esteem changes during their undergraduate years. This two-factor model of self-esteem recognized that self-esteem includes factors of both worthiness and competence. Worthiness involves one’s feelings of their self-worth and competence involves one’s assessment of their skills. When these two are balanced and positive, one feels happy, mentally healthy, and authentic. In Hippler’s cross-sectional research, 131 undergraduate and professional women, majoring and working in computer science and related fields completed a computing experience survey based on the Rosenberg Self-esteem Scale. Using a two-factor model of self-esteem, Hippler demonstrated that these two dimensions are out of balance - although some women may feel competent about their skills in computing, they have doubts about whether they or their work is worthy. This work suggests that intervention programs focused on developing competence along with worthiness could help retain women in computing. The good news is that there is over 100 years of self-esteem research to draw upon. Programs such as Mruk’s Competence and Worthiness Training (CWT) can be used to help students find sources of competence and worthiness that promote a healthy self-esteem.

Using the CWT as a foundation, a 45-minute workshop was developed by this author and delivered to 48 college students and professors attending a state-wide, women in computing conference. The workshop provided attendees with an awareness of how self-esteem works and its connection to academic success, validation of their feelings, and ideas for how to take positive action to improve their self-esteem. This workshop was one of five different activities that students had the option of attending. A quarter of the students at the conference choose to attend this workshop serving as an indication that the topic is relevant. Furthermore, the wide distribution of attendees indicate that this sort of workshop was perceived to be valuable to all students regardless of academic status (freshman, sophomore, junior, senior, graduate) as well as to faculty.

Upon arrival, the students received a workbook to use during the workshop. The workshop explained authentic self-esteem, why authentic self-esteem matters, and the importance of authentic self-esteem. The workshop then provided time for the attendees to identify and share general sources of self-esteem and sources of self-esteem found within computing. The workshop wrapped up with suggestions to create an action plan, journal about positive experiences, and set reminders to reflect.

Thirty-six attendees provided feedback on the workshop: freshmen (N=5), sophomore (N=6), junior (N=13), senior (N=8), graduate students (N=2), professional (N=2). Students and faculty confirmed the importance of the workshop by indicating that they liked the workshop, found it valuable, and wanted more (average scores of 3.5 on a 4-point Likert-scale). Additionally, half wanted a longer workshop and half asked for more discussion. Moreover, half of the survey responders are willing to talk more about the workshop through a follow up interview. (The interviews will take place prior to the fall Technology, Mind, and Society Conference and will be transcribed, coded, and reported at that time).

Although the author plans to refine and deliver this workshop to other student groups, she hopes to also help faculty. Teaching computer science comes with challenges. First, students often do not

have the development and/or cognitive style to deal with emotions (their own or others') that make up self-esteem. Second, faculty who prefer to use logic over emotion may be more problem-focused than emotional-focused. This can impact what they see what they do not see, what they ignore, and how much flexibility or rigidity is considered in situations. And third, computer science has been often characterized as having a defensive climate with a focus on skills and logic over emotions. This author wonders if the chilly (analytical) nature of the major could benefit from a warmer (emotional) perspective and whether the discipline could benefit from learning to see things from the other side such as thinking about computing from the human side. Therefore, a next step in this research might be to create a self-esteem workshop for faculty to help attendees appreciate the feelings that women in computing (and maybe men too) have and learn how to focus conversations on (1) positive experiences and (2) things the students can control.

*Rachelle Hippler (Baldwin Wallace University)*

## Mobile Computing / Technology

### I-1

#### **A Smartphone Ecological Momentary Assessment 'App' for Collecting Real-Time Data on Posttraumatic Stress Injuries in Firefighters**

Public safety personnel, including firefighters and first responders, can be exposed to highly disturbing or threatening events, such as badly injured people, deceased children, and highly distressed individuals. On average, firefighters attend to approximately eight different potentially traumatic events at work annually. Mental health problems, labelled as posttraumatic stress injuries (PTSI), can result from those exposures and may include anxiety, depression, substance abuse, and post-traumatic stress disorder (PTSD). The National Institute of Health and Care Excellence (NICE) suggests that within the first month following the traumatic event, we should repeatedly assess the employee exposed to a traumatic event in order to detect early development of symptoms and help manage these difficulties before aggravation. New technologies, such as mobile device applications, may be a useful way to complete repeated assessments in real time. The proposed research is a longitudinal study that intends to apply NICE's recommendation by monitoring PTSI symptoms in real time (via a mobile application) among trauma-exposed firefighters and to study the trajectory of PTSI and associated factors over three months. Two hundred male firefighters will be recruited throughout all fire stations in Montreal, Canada. Recruited participants will be asked to complete several questionnaires via the mobile application every two weeks over a 3-month period. The proposed study will help to improve clinical interventions addressed to firefighters exposed to traumatic events by identifying the presence of risk factors for PTSI in real time. The use of a mobile application allows for good cost-efficacy in terms of the delivery of interventions.

*Stephane Guay (University of Montreal)*

### I-2

#### **Boredom Proneness Mediates Relations Between Depression Severity and Interruptions in Daily Activities from Smartphone Notifications**

Interruptive notifications are information alerts that intend to actively draw away one's attention in order to inform of a new event or information. Several studies have examined adverse effects of interruptive computer notifications on work or school performance. However, little research has examined interruptive smartphone notifications, or individual differences that may account for adaptive vs. maladaptive effects of interruptive notifications. Relevant to this topic is research on problematic smartphone use (PSU), as prior research has found that habitual checking for smartphone notifications is related to PSU severity. We used flow theory and the Interaction of Person-Affect-Cognition-Execution (I-PACE) model to conceptualize personality and mental health variables impacting the adverse effects of interruptive smartphone notifications. We recruited 316 college students for a web-based survey, inquiring about 20 daily activities in which individuals participated (e.g., socializing, sleeping, doing schoolwork, etc.) and frequency of responding to interruptive smartphone notifications during these activities, depression (DASS-21 Depression Scale), anxiety (DASS-21 Anxiety Scale), proneness to boredom (Boredom Proneness Scale-Short Form), and emotional distress tolerance (Distress Tolerance Scale). We tested a structural equation model using weighted least squares with a mean- and variance-adjusted chi-square. We tested mediation using the Delta method, computing cross-products of direct effects with 1000 non-parametric bootstrapped replications. Depression and anxiety (predisposition variables) were specified as predictors of boredom proneness and distress tolerance (cognitive and affective response variables), in turn serving as predictors of activity interruptions from smartphones, adjusting for age and sex. Boredom proneness significantly predicted activity interruptions (standardized beta = .31, SE = .05). Furthermore, depression severity was related to boredom proneness (beta = .99, SE = .31), and inversely related to distress tolerance (beta = -.43, SE = .15). Boredom proneness mediated relations between depression severity and activity interruptions (beta = .37, SE = .11). This study suggests that important mental health variables (i.e., depression and anxiety) may be related to increased interruptions in daily activities from smartphone notifications, and cognitive and emotion processes such as boredom proneness and distress tolerance may play a role.

*Jon Elhai (University of Toledo)*

## Neuroscience

### J-1

#### **The Therapeutic iPad: Technology in Neuropsychology**

Statement of the problem: Technology is a force that continues to grow and expand in many areas of human life. People tend to use technology in their daily life, and this appears to impact their interaction with themselves and with others. While technology has advanced in several helping professions, such as medicine and physical rehabilitation, technology continues to be absent in clinical psychology, more specifically, in neuropsychology. Neuropsychology focuses on the assessment of cognitive functions of the brain and their development throughout the lifespan. There exist many different

tests and batteries that assess a range of psychological constructs. Many of these tests have been applied in clinical practice for several decades and continue to be updated. Some of these tests possess computer programs that help psychologists score and interpret the scores obtained. This appears to create a faster process both for patients and for psychologists. There has been a recent movement that attempts to eradicate paper and pencil forms. Companies such as Pearson have developed computerized applications that intend to make the psychological assessments faster and user-friendly. However, this innovation has not been well received in the neuropsychological field. One of the problems with these types of applications are the norms for the population and the scores that patients receive, as these do not translate to scores obtained in paper and pencil tests. There appears to be controversial opinions regarding the use of technology in neuropsychological assessment. The current poster intends to address this new area of clinical practice and focus on the different problematic areas that may come up for patients and psychologists.

**Procedures:** The current poster will conduct a literature review on the use of technology in neuropsychological assessment. It appears that using technological devices, such as computers and tablets, has become a highly debated topic in recent years. There appears to be a strong side that explains the benefits of using technology to help neuropsychologists assess for cognitive deficits. However, there is also another side that condemns the use of technology in assisting neuropsychological assessment. In doing a literature review, the poster will generate ideas and illustrate the gaps in the literature. Additionally, the poster will present comments and perceptions of neuropsychologists who currently conduct neuropsychological assessment. By including this information, the poster seeks to have different views on the use of technology in assessment and create a discussion regarding this controversial topic.

**Analyses:** This poster will summarize current and relevant research. In doing so, the poster will present the different sides that exist on this topic as the use of technology continues to increase in daily life and in other health related professions. The poster will also analyze qualitative responses from neuropsychologists in order to explore different perspectives and opinions regarding the use of technology in clinical practice.

**Results:** The poster intends to show different problems and solutions to the use of technology in neuropsychology. It is expected that the results allow for further discussion and possible implications in clinical practice as it relates to expert opinions.

**Practical implications:** The results from this investigation may provide important considerations when using technology in clinical practice. By using a literature review, the results will allow the readers to focus on consistent thoughts and opinions in order to promote an open discussion. Additionally, the literature review will offer a basis for how the changes in technology are affecting the field of neuropsychology and how these changes are being received by the professionals. Furthermore, it is expected that technology in this area continue to grow, thus it is important to consider neuropsychologists' perceptions of their own field and the impact of technology.

**Conclusions:** The poster will attempt to answer questions related to the benefits and the disadvantages of technology-based neuropsychological assessments. The poster will promote a discussion regarding the role of human interaction in assessment. Finally, the poster will present ideas for future research and the directions of the field.

*Sebastian Del Corral Winder (University of Indianapolis)*

## J-2

### **Brain Lateralization and Use of Positron Emission Tomography (PET) on Emotional Processing of Clinically Diagnosed Anxious and Depressed Male Patients**

Lateralization is the idea that the two halves of the brain's cerebral cortex, left and right hemispheres, execute different functions. This study employed small-N design involving 4 male patients (2 anxious with left and right dominance) and (2 depressed with left and right dominance) who were clinically diagnosed in mental facilities. The emotional processing was done during the PET scan where the framed pictures were mounted at the PET machine above the head region of the patient. The brain scanning lasted for 10 minutes. A post-inquiry was conducted using the Self-Assessment Manikin (SAM) of Bradley and Lang (2007). PET scan images revealed that both the anxious left and right hemispheres dominant had hypermetabolism in the right sensorimotor cortex during the emotional processing. The depressed left hemisphere dominant patient revealed hypermetabolism in the left midfrontal cortex and left associative visual cortex while the depressed right hemisphere dominant patient revealed hypermetabolism in the right primary visual cortex. Using the SAM, anxious right hemisphere patient revealed happy emotion with the pleasant picture and felt in control with the unpleasant picture while the anxious left hemisphere dominant revealed neither happy nor unhappy with the pleasant picture but felt dominated with the unpleasant picture. On the other hand, both left and right hemispheres dominant depressed patients revealed happy feelings with pleasant picture and felt dominated with the unpleasant picture. Considering the brain dominance of the patients, it implied that anxious patients processed emotions in the right region of the brain regardless of dominance implicating that the right hemisphere processed emotions regardless of valence. On the other hand, depressed patients processed their emotions based on brain dominance.

*Hector Perez (St. Scholastica's College - Manila, Philippines)*

## J-3

### **Development of a Measure for Science Literacy Leveraging Neuroscience and Machine Learning**

**Problem:** According to the Trends in International Mathematics and Science Study (TIMSS), the US science scores have increased between 1995 and 2015 for both 4th and 8th grade students (Provasnik et al., 2015) and although the US is keeping up pace with the other nations, the faster growth in science and technology of other nations may make it more difficult for the United States to stay competitive without a focus on how science pedagogy. However, the results of these examinations do not fully get to the idea of scientific literacy. Currently, only 28% of adults in the US are considered scientifically accurate and this lack of achievement of scientific literacy is primarily due to the enormous amount of variation in educational standards within the United States, as well as failure by educational research to move beyond a product oriented focus to understanding the learning process in science.

An integral part of learning science involves visual and spatial thinking since complex and real-world phenomena are often expressed in visual, symbolic, and concrete modes (Ramadas, 2009). The primary barrier to spatial thinking and visual literacy in STEM fields is representational competence, which includes the

ability to generate, transform, analyze and explain representations, as opposed to generic spatial ability (Stieff et al., 2018). Although the relationship is known between the foundational visual literacy and the domain specific science literacy, science literacy as a function of science learning is still not well understood. Moreover, the need for a more reliable measure is necessary to design resources which enhance the fundamental visuospatial cognitive processes behind scientific literacy.

In order to develop pedagogies to support the improvement of students' representational competence, it was important to first identify and assess the necessary visualization skills to process these representations which requires the development of an instrument to measure visual literacy. With a measure of visual literacy, schools, teachers and curriculum designers can target the individual skills necessary to improve students' visual literacy, thereby increasing science academic achievement. One such measure was developed in 2016 by Mnguni et al. Mnguni's measure required the provisioning of a series of probes single dimension visually based questions which required content knowledge to answer. Since Mnguni's self-report study determined that a student's level of visual literacy can be determined by visualization probes, unbiased psychophysiological measurements should be able to be used as a visual literacy assessment as well.

Therefore, this project details the development of an artificial neural network capable of measuring science literacy using functional Near-Infrared Spectroscopy (fNIR) data previously collected by Project LENS, a Science of Learning Collaborative Network (SLCN) of scholars of STEM Education, neuroscience, cognitive psychology, computer science and educational technology and measurement from 3 universities working to advance research about learning through integrative and empirical research (NSF award 1540888).

**Procedures:** Brain blood deoxygenation data from fNIRsoft was exported as an excel file with tasks categorized as dash for 2D Wedge and Dash models and BL for 3D Stick and Ball models. Problem complexity data was kept in an Excel workbook separated by participant (ID), containing information for both Ball and Dash tasks. After labeling both datasets and changing strings to numbers for analysis, the spreadsheets with measurement data and complexity data were uploaded to RapidMiner Studio's TurboPrep and merged (outer join) using the ID number as the Join Key. Then, using RapidMiner's AutoModel feature, several forecasting algorithms were compared in order to identify the most accurate predictive model for the data.

**Results & Analysis:** The Gradient Boosted Trees (GBT) model had a higher overall classification accuracy (Figure 1) and lower classification error (Figure 2) than each of the other predictive models.

Project LENS utilized Mental Rotation tasks, a visualization task of average difficulty, to assess a student's visual literacy. Two types of representations of molecular structures were used; 3D Ball and Stick models (Ball) and 2D Wedge and Dash models (Dash). Ball tasks were completed 1520 times; the participant was correct 1023 (67.3%) times. Of the 1439 times in which Dash tasks were completed, 1062 times (73.8%) the participant was correct.

The Gradient Boosted Trees artificial neural network consisted of 140 trees with a maximum depth of 7 branches had an accuracy of 99.7% found that the biggest predictors to a successful mental rotation are the individual problem number of the 160 mental rotation items performed by each participant (80 Ball and 80 Dash), the Response Time and fNIR optode #16, located along the right prefrontal cortex which plays a vital role in processing visuospatial

working memory (VSWM) (Suzuki et al., 2018) and episodic memory retrieval (Henson, Shallice, & Dolan, 1999).

With an unbiased measurement of science literacy provided by autonomic blood deoxygenation results and an artificial neural network for analysis, teachers will have the ability to create more targeted classroom resources to help improve student visual literacy, thereby improving their science literacy.

*Amanda Kavner (SUNY Buffalo)*

## Public Safety

### K-1

#### Hijacking Mitigation Techniques: Should Pilots Act?

**Background:** Prior to the terrorist attacks of September 11, 2001, security measures within airports were significantly more relaxed (Blalock, Kadiyali, & Simon, 2007). After 9/11, airport security received a massive overhaul in an effort to ensure passenger safety, restore passenger confidence, and stop future hijackings. Many new policies involved tasking Security Administration (TSA) employees with visually searching for illegal items (Biggs, Cain, Clark, Darling, & Mitroff, 2013; Blalock et al., 2017; McCarley, 2009). Unfortunately, humans often perform poorly when tasked with searching for multiple targets/stimuli, particularly when the probability of the target/stimulus presence is rare (Menner, Donnelly, Godwin, & Cave, 2010; Mitroff & Biggs, 2014). For example, only about 1%, or lower, of luggage going through airport security screening actually contains illegal items, making it an extremely rare item with a high chance of employees missing it (Mitroff & Biggs, 2014; Wolfe, Horowitz, & Kenner, 2005). With such a high failure rate, it's critical to identify additional security measures that can be implemented to stop a potential hijacking. Therefore, the purpose of this study is to provide participants with different hijacking mitigation strategies and identify which ones they support or oppose.

**Methods:** Two hundred and eight participants were recruited via Amazon's Mechanical Turk © (MTurk) from the United States and were presented with a survey containing one of two hypothetical scenarios. Participants read, "Janice/Thomas Hollins is a commercial pilot who is a captain at a major airline. S/he is flying en route between two major cities and during the flight, a hijacker attempts to take control of the passenger section of the airplane." Participants in the female pilot condition read the scenario with Janice's name (and female gendered pronouns) and participants in the male pilot condition read the scenario with Thomas' name (and male gendered pronouns). Next, participants were shown a list of different hijacker mitigation strategies and asked to assume that the pilot was fully trained in each strategy. Participants used a Likert-type scale to indicate how strongly they agreed or disagreed to the pilot utilizing each strategy. The different strategies were as follows:

The pilot should be allowed to... 1. Come out of the cockpit with a handgun and shoot the hijacker; 2. Come out of the cockpit with a Taser, which is an electroshock weapon, and tase the hijacker; 3. Depressurize the entire cabin area, causing everyone to pass out, so that the hijacker can be detained; 4. Come out of the cockpit with a shotgun, which fires a non-lethal bean bag round, and shoot the hijacker; 5. Come out of the cockpit with a tactical flashlight, and blind and stun the hijacker.

After participants finished marking their responses, they provided an answer to the prompt, “Briefly describe how you made your decision about what the pilot should or should not do.” Finally, participants answered basic demographic questions and then they were debriefed, compensated, and dismissed.

Results: The purpose of this research was to determine which hijacking mitigation technique participants would be most comfortable with a pilot utilizing to stop, or hinder, a hijacker’s efforts. A two-way analysis of variance was conducted on the data using Pilot Gender and Mitigation Tactic as the factors. There was a significant main effect of Mitigation Tactic,  $F(4, 784) = 26.07, p < .001$ , partial eta squared = .12, with participants reporting the greatest amount of support for pilots utilizing a taser and the least amount of support for pilots depressurizing the cabin (all  $ps < .01$ ). There was not a significant main effect for Pilot Gender nor was there an interaction between the two factors. Furthermore, a thematic analysis was conducted on the qualitative question, revealing that, in general, participants did not want the pilot to leave the cockpit or utilize any strategy that could indirectly innocent passengers. Participants also suggested that there should be security personnel onboard to help neutralize the threat rather than the pilot being in charge of maintaining aircraft control and mitigating a hijacking.

Conclusions: With increased security measures, airports are striving to improve safety and decrease the likelihood of hostile passenger behavior, such as hijackings. Unfortunately, current security measures often miss illegal, potentially dangerous, items that a hijacker could use. However, additional security measures could be implemented, such as allowing pilots the authority to take defensive action against a hijacker. Participants indicated the strongest amount of agreement for a pilot (or some other security personnel) using a taser to incapacitate a hijacker and the least amount of support for depressurizing the cabin, which could potentially harm innocent passengers. While a lot of research has explored factors addressing high-workload cognitive tasks, such as visual searches, research has neglected to address what happens when a person inevitably misses a target/stimulus in their visual search, particularly when a failure could cause a monumental detriment to safety and security.

*Mattie Milner (Embry-Riddle Aeronautical University)*

## Social Media

### L-1

#### The Impact of Mindsets about Social Media on Psychological Well-Being

We present theory and data informing the development, validation, and predictive validity of the 20-item Social Media Mindsets Scale and the SMM framework on media use and well-being, with subscales of valence (“Social media is enhancing” vs. “harmful”) and perceived user agency (“Social media serves users” vs. “controls users”). The growing body of research on mindset has found that beliefs can shape key psychological, behavioral, and physiological responses. The way that individuals think about concepts like intelligence and stress can change how it impacts them (Dweck, 2006; Crum, 2013; Zion, 2018; Mangels et al. 2006). Social media usage has been found to have both positive and negative effects on well-being, including social connection, life satisfaction, affective experience,

and the experience of depression and anxiety (Kross et al., 2013; Verduyn et al., 2015; Burke et al., 2010; Ellison et al., 2011; Hancock et al., forthcoming). Conflicting results for the same well-being variables are widespread in the literature; for example, increased use of social media has been found to both increase and decrease loneliness across 60 studies (Huang et al., 2010; Hancock et al., 2018). These mixed positive and negative results emerge from experimental, correlational, longitudinal, and survey-based empirical measures, indicating that these dissonant findings are not a byproduct of methodology variance, but rather a reflection the divergent nature of social media’s influence on well-being. However, no research has yet explored the role of mindsets—an individual’s set of beliefs and expectations—on this relationship. By building on existing literature in social psychology, media psychology, and mindset, we developed a framework and scale for measuring social media mindsets along the dimensions of perceived valence (“Social media is enhancing” vs. “Social media is harmful”) and perceived user agency (“Social media is a tool controlled by users” vs. “Social media exerts control and influence over users”). We used a mixed-methods approach across three studies to qualitatively assess user mindsets about social media through interviews and to quantitatively assess the reliability, validity, and predictive power of the 20-item Social Media Mindsets Scale. Our samples included college students ( $n = 250$ ) and mTURKers ( $n = 300$ ). We found the scale to have good internal consistency ( $\alpha = 0.88$ ) and normal distribution. In addition, the SMMS predicts the well-being outcomes of loneliness, relational well-being, anxiety, and satisfaction with life significantly and demonstrates discriminant validity from existing measures of social media usage (Facebook Intensity Scale, self-reported usage hours, passive vs. active use, network strength). Together, these studies suggest that social media mindsets is a distinct and meaningful variable in understanding the relationship between social media usage and well-being. The present work is significant because understanding how mindsets about social media interact with usage serves as a potential mechanism of how media use impacts users’ health and well-being. In addition, the malleability of mindset also presents a unique opportunity to intervene wisely at the level of belief to improve participants’ outcomes.

*Angela Lee (Stanford University)*

### L-2

#### The Instrumentality of Emotion on Social Media: Self-Relevance as the Moderator in Emotional Sharing on Twitter

Emotion is believed to play an important role in shaping sharing behavior on social media. However, the underlying mechanism is not clear yet. Existing findings are mixed on whether the intensity of online emotion sharing increases during social transmission due to echo chamber effects or it decreases as calming effects through making a closure. We adopt a linear mixed model to analyze a large-scale dataset extracted from Twitter discussions about 6 topics which varied on morality and self-relevance ( $n=142,703$ ). The result that in moral issue discussions, the difference of emotion intensity between original tweets and retweets were positively associated with more responses and further shares. The online mechanism of changing emotion intensity could be that individuals are actually rewarded by upregulating or downregulating emotions through social interaction.

However, consistent with previous findings (Brady, et.al, 2017; Goldenberg, Grossm Gracia, 2018; Fan, et.al, 2019), the emotion changing pattern varied across the discussions and the mechanism was left unanswered. The further analysis of a reference discussion group, a casual discussion on puppy showed the above emotion changing rewards no longer influenced the shares or responses. Compared these groups, the results revealed that self-relevance, the relative frequency of mentioning “we”/“I” in tweets, moderated the association between emotional intensity and social interaction. In highly self-related discussions (abortion/gun control/same-sex marriage), increasing emotion intensity resulted in more shares whereas in low self-related discussions (climate change/scientific progress/puppy), increasing emotion intensity resulted in fewer shares or showed no effects. This may be because of individuals’ selective attention to self-relevant information which makes them emotionally ignore objective facts.

*Amber Xuqian Chen (The Chinese University of Hong Kong)*

## KEYNOTE ADDRESS

9:15–10:30 A.M.

Independence Ballroom

## Stealth Assessment — What, Why, and How?

Games can be powerful vehicles to support learning, but this hinges on getting the assessment part right. In the past several years, we have designed, developed, and evaluated a number of stealth assessments in games to see: (a) if they provide valid and reliable estimates of students' developing competencies (e.g., in the areas of qualitative physics understanding, creativity, and persistence); (b) if students can actually learn anything as a function of gameplay; (c) the added value of inserting engaging learning supports (cognitive and affective) into the mix; and (d) if the games are still fun with the embedded assessments and supports.

This presentation will cover the topic of stealth assessment in games to measure and support important 21st century competencies. Dr. Shute will cover why it's important, what it is, and how to develop/accomplish it. She'll also provide numerous examples and videos in the context of a game called Physics Playground.



### VALERIE SHUTE, PhD

**Mack & Effie Campbell Tyner Endowed  
Professor in Education**

*Department of Educational Psychology  
and Learning Systems, Florida State University*



10:30–11:00 A.M.

Independence Foyer

## Networking Break

---

### Meet the Editor of APA's NEW journal *Technology, Mind, and Behavior*

APA's new open access journal, *Technology, Mind, and Behavior*, will publish innovative research on human-technology interaction. The journal will publish articles in various areas, including basic research on technology and the human experience, foundations of technology design, and the broader implications of human-technology interaction.

Stop by the APA Publishing networking break to chat with founding Editor, Dr. Danielle McNamara and learn more about the journal, including how to submit your manuscript.

11:00 A.M.–12:15 P.M.

## Concurrent Sessions 4

---

Independence D

### Challenges and Opportunities for Using Big Data

#### PAPER SESSION

#### Open-Sourcing AI Development Through Machine Learning Competitions: Vision, Process, and Learnings

Advancements are being made in computer science and artificial intelligence with increasing speed. Meanwhile, academic processes are hard-pressed to adapt the slow modalities of peer-review journals to assess the plethora of methods available to keep pace. We propose that a competition built on open source principles can more rapidly prototype solutions, creating a transparent benchmark for future models, methods, and algorithms. While our example is in the field of industrial-organizational psychology, the competition model we showcase can be applied to any field. Our presentation will illuminate the process and tools of conducting such an event, with attendees walking away with the knowledge of how to conduct a world-class machine learning competition.

Diversity. The 2nd annual SIOP (Society for Industrial and Organizational Psychology) machine learning competition was modeled after the world class machine learning competitions hosted on kaggle.com (Valentino-DeVries, 2011) and the cha-learn competition (Ponce-López et al., 2016). Both of these modalities open a) participation, b) the data, and c) the winning code to the community. Opening participation to the widest audience possible ensures that diverse solutions are put forth. Diversity of approaches offers a competitive advantage that preconceived hierarchies can unintentionally limit. Winning solutions are not biased to a certain country, ethnicity, or any other background. No one is prohibited from participating or winning, thus it is a true meritocracy. Yet, scaling a competition to an unlimited number of participants has its own challenges which require automation to overcome.

Automation. Crucial to our success was automation of all systems ranging from gathering participation to sharing data and analyzing teams submissions. For gathering participation we used Google forms. This provided us a reasonable way to track signups, automate the sharing of data to participants, share competitions rules, and gather demographic data. For hosting the competition we partnered with a team of computer science students at Georgia Institute of

Technology and used their technology, EvalAI (<https://evalai.cloudcv.org/>). We worked with them in the open source program they built to provide a portal for participants to build teams, submit, track, and evaluate results, and view a live public leaderboard. The live leaderboard encourages constant improvement of solutions, as it automates the performance feedback loop. Scripting tasks on our side were limited to building out the content of the website and the results evaluation script.

Benchmarks. Another important goal of our competition was to create benchmarks from the data and the solutions/methods, specifically for solving unique problems facing I-O psychologists in the workplace (e.g., personnel selection). Data commonly used in the machine learning communities (e.g., reddit) may not be relevant or suitable for solving problems in other fields. The field of I-O psychology has years of expertise in measurement of human traits. Leveraging the I-O subject matter expertise and knowledge of data requirements for ML applications, we were able to create a benchmark dataset that is not only suitable for this competition but also available to use for future research and applications.

Data. Sharing human measurement data can be difficult. For a majority of companies human resource data is some of their most private and secure data. To open this competition up to the most diverse set of participants, we had to think creatively how to gather data that was high fidelity yet did not fall under a confidentiality agreement. Also, as our field has virtually zero peer reviewed articles on the power of the more advanced methods of machine learning, deep learning, we wanted a data set that lent itself to documenting the power of deep learning. To meet these diverse needs, we set out to collect data via Mechanical Turk. To make it relevant to our field of work psychology, we asked participants a set of 5 structured job-relevant situations that were built off of situational judgement items of personality (Oostrom, de Vries, & de Wit, 2019). We then ask study participants to fill out a well developed big five multiple choice inventory (Soto & John, 2017). The primary task for competition participants was to predict personality from the open ended texts responses.

Results and Learnings. To transform analytical methods in a given field, algorithms must be made open and transparent. From our competition, winning teams' source code that will replicate results is made permanently available alongside links to freely download the data on Github. Prior to the closing of the event we will be collecting additional data on the 80 plus teams that participated answer what methods and programming languages they used. During our presentation we will illustrate what learnings the field can make from this competition, process, and open algorithms. Lastly, based on this trajectory we will share informed hypotheses on where the field of assessment is heading and how it will get there.

*Isaac Thompson (Shaker)*

### **Barriers and Challenges to Big Data Uptake in Work and Organizational Psychology: An Analytical Framework at Individual, Team, and Organizational Level**

Big Data, generally defined as data being high in volume, variety, and velocity (Guzzo, Fink, King, Tonidandel, & Landis, 2015), is a key technology of the Fourth Industrial Revolution (Badri, Boudreau-Trudel, & Souissi, 2018). Its application areas (Table 1) include human resource management (Gelbard, Ramon-Gonen, Carmeli, Bittmann, & Talyansky, 2018), customer experience (Yom-Tov, Rafaeli, Altman, & Ashtar, 2017), organizational design (Clegg et al., 2017), research methods (Braun, Kuljanin, & DeShon, 2018), and workplace health and safety (Rivas et al., 2011).

Benefits deriving from BD involve both academia and practice. "The massive volume of data that now covers a wide variety of human behaviors offers researchers in psychology an unprecedented opportunity to conduct innovative and theory- and data-driven field research" (Chen & Wojcik, 2016; p. 458). As well, BD methods can predict fundamental HR phenomena like performance, engagement, leadership, workplace dynamics, organizational learning and development, and allow for real-time, less costly, and more reliable assessments of human factor patterns (Gelbard et al., 2018).

In light of such positive impacts, it seems important to detect which factors are inhibiting the access to, adoption and use of BD technology in the Work and Organizational Psychology domain. Indeed, this knowledge would inform solutions designed to mitigate the influence of such hindering factors and promote BD benefits exploitation.

In the current study, we pursued this aim by adopting the three-level analytical framework (individual, team, organization) provided by the discipline itself. In so doing, we aimed at reconceptualizing peculiar issues of the BD realm through the application of fitting constructs typically belonging to the WOP scientific field.

To this end, we conducted a literature review on Web of Science, with the scope of answering the following research question: What are the barriers to the uptake of Big Data in Work and Organizational Psychology research and practice at individual, team, and organizational level of analysis?

Without refining for timespan nor WoS categories nor document type nor indexes, we searched for the query: TOPIC("big data" OR "data mining") AND ("psychology" OR "workplace\*" OR "future workplace\*" OR "future of workplace\*" OR "hrm" OR "human resource\* management"). We retrieved 450 results, of which 36 were selected since considered relevant to our RQ. Thirteen additional references were included into the review based on a snowball sampling procedure. Ultimately, a total of 49 references was

reviewed, spanning from 2011 to 2019, therefore indicating Big Data as a trending topic in last decade.

Based on our expert judgement, we clustered three sets of barriers, according to the number of adopted levels of analysis.

At individual level, we grouped the two overarching factors "lack of skills" and "negative attitudes". "Skills" referred both to hard/technical (expertise in advanced statistical techniques, software programming skills, IT skills, computational skills) and soft/cross-cutting (analytical thinking, interdisciplinary skills, and communication skills). "Negative attitudes", in turn, differentiated in: "Concerns about confidentiality or anonymity violation"; "Skepticism towards BD methods effectiveness"; "Motivation and openness to change". Interestingly consistent to the latter result, a classical literature stream in Psychology (Bagozzi, 2007; Venkatesh, Morris, Davis, & Davis, 2003) shows attitudes' role as technology uptake antecedents. Further, privacy concerns might be interpreted as standing for a low degree of trust towards BD technology which, re-elaborating Rousseau, Sitkin, Burt, and Camerer (1998), may be defined as a psychological state comprising the intention to accept vulnerability based upon positive expectation of BD behavior (i.e., "I expect that BD will not violate my privacy").

At team level, we identified "ineffective leadership", "teamwork characteristics", and "unavailability of transdisciplinary teams". Leaders may struggle in taking all necessary decisions to change organizations toward the new BD era, teams not always adopt right collaboration modes for achieving such goal (Buchanan et al., 2017), and multiple expertise teams are often difficult to find (Cheung & Jak, 2016).

Finally, at organizational level, major barriers to BD adoption and use were "lack of technological infrastructure", "lack of organizational support towards innovation", "organizational culture", and "organizational hindering features". Particularly, since HR practices are good company culture's indicators (Aycan, 2005), we considered "organizational culture" as being reflected by "lack of internal communication" ("silo mentalities"; Angrave, Charwood, Kirkpatrick, Lawrence, & Stuart, 2016) and "old-fashioned training mode" (Deng & Cao, 2018). Such results are in line with WOP contributions stating that certain environmental/situational determinants, such as availability of technological facilities (Shattow, 1996), social support (West, 2002), and corporate culture (Martins & Terblanche, 2003), are important resources acting as sustaining mechanisms for successfully implementing innovative processes.

Theoretically, the current study suggests a framework for analyzing Big Data issues from the viewpoint of Work and Organizational Psychology. Practically, our results provide concrete insights related to the design of education and training programs for both universities and companies (e.g., upskilling, team management, leadership development), as well as to the building of data management plans (e.g., aimed to protecting users' privacy).

*Davide Giusino (University of Bologna)*

### **Combining the Data and Information Sciences to Organize Psychological Knowledge**

I use the term 'data sciences' to refer to the collection of tools such as machine learning that support bottom-up reasoning from data. Machine learning is having an increasing influence on our lives as indicated in the book *The Master Algorithm: How the Quest for the Ultimate Learning Machine will Remake our World* (Domingos, 2015). The author, a Professor of Computer Science

at the University of Washington, describes five ‘tribes’ of machine learning: the Analogizers, Bayesians, Connectionists, Evolutionaries, and Symbolists. His book gives readers a clear introduction to these methods and their applications.

The Master Algorithm encourages thinking about how combining these methods can make them more effective and challenges readers to help achieve this goal. Psychologists have done extensive work on developing these methods, which I describe in my article ‘Building Bridges Between AI and Cognitive Psychology’ in *AI Magazine* (Reed, in press). I conclude by discussing three benefits of building bridges: solving joint computational problems, using computational models in AI as theoretical models in Cognitive Psychology, and facilitating the interactions between people and machines.

I use the term ‘information sciences’ to refer to the collection of computer tools that support top-down reasoning to organize data. The hierarchical representation of knowledge plays a central role in the creation of formal ontologies by computer scientists. Ontologies consist of a hierarchy in which the top node is partitioned into smaller categories that compose the hierarchy. Formal ontologies add a logic that enables deductive reasoning from statements expressed in a computer language.

I was surprised that, with only a few exceptions (Yarkoni, 2012), there had not been more advocates of the importance of ontologies for Psychology. I began to understand the reason when I began my own explorations of this field. There were few tutorial articles on ontologies and many of the contributions were buried in the proceedings of engineering conferences. Then a book, *Ontology: A Practical Guide* (Pease, 2011), came to my rescue. The author, Adam Pease, is the technical editor of the Suggested Upper Merged Ontology (SUMO), a large ontology owned by the Institute of Electrical and Electronics Engineers (IEEE). Pease and I subsequently collaborated on the article ‘A Framework for Constructing Cognition Ontologies using WordNet, FrameNet, and SUMO’ to describe Information Science tools to cognitive scientists (Reed & Pease, 2015).

Adam Pease and I extended this overview in our article ‘Reasoning From Imperfect Knowledge’ (Reed & Pease, 2017). Although both the Information and the Cognitive Sciences contribute to reasoning from knowledge, there are few cross-references between these two domains. One of the reasons is that they have different missions. Psychology studies how people reason and the Information Sciences provide technology tools to help people reason. However, both people and computers face the same challenges when confronted with ambiguous, conditional, contradictory, fragmented, inert, misclassified, or uncertain knowledge. Our article provides examples of how both people and computers face these challenges.

A grand challenge for Psychology is the NIMH RDoC initiative (Insel & Cuthbert, 2015) to establish a biological basis for mental illness by linking units-of-analysis with functions-of-behavior.

A major problem for filling in the cells of Table 1 is to represent such large, complex, and multi-scale knowledge in a manner that can support computational analysis, including query answering. The Semanticscience Integrated Ontology (Dumontier et al., 2014) offers a scaffold in which its initial emphasis on molecular biology could be extended upward to neurological and psychological functions. I collaborated with the lead developer of the Semanticscience Integrated Ontology, Michel Dumontier, to develop examples of this extension within RDoC’s units-of-analysis and functions-of-behavior dimensions (Reed & Dumontier, 2019).

In conclusion, we need to combine the Data and Information Sciences to organize psychological knowledge. WatsonPaths, which

expands on the Watson question-answering system that became famous on Jeopardy, provides an impressive application to medicine (Lally et al., 2017). The project presents a patient summary and asks for the most likely diagnosis or most appropriate treatment. Guiding the diagnosis is a model of entity types and relations. An application of these ontological relations connects input symptoms to hypotheses with probabilities as mediators. The big data revolution should provide large amounts of data for estimating probabilities to support probabilistic deductive reasoning by combining the Data and Information Sciences (Poldrack & Yarkoni, 2016).

*Stephen Reed (San Diego State University)*

### How AI Is Revolutionizing Psychometrics

Artificial Intelligence (AI), and in particular deep learning (DL), has revolutionized predictive modeling. These techniques, coupled with the ability to capture and store vast amounts of data, are opening exciting new avenues of research, offering psychologists the potential to make important and needed contributions to understanding human behavior.

For the potential of AI in psychology to be realized, we must consider the power of analytical techniques that are foreign to many in the field of psychometrics, and for which we have little training. Our ability to do so—to unlearn and relearn—will determine our relevance to scientific progress in the age of AI. This paper will present the argument that AI is transformative for psychometrics, and requires wholesale rethinking of what are considered fundamental tenets of measurement and analysis.

What exactly is the power of deep learning and how is it transformative over traditional psychometric techniques? At a very basic level, deep learning occurs when neural networks are stacked on top of each other. In contrast to traditional analysis techniques, this is loosely analogous to how the human brain operates. And so it follows that Deep Learning is capable of processing larger amounts of less structured information than common techniques, and doing so with an extremely high level of accuracy. This is analogous to how the human brain quickly processes vast amounts of sensory data to derive often unconscious conclusions, such as when you “get a bad feeling” about someone or meet a new person and “thin slice” (Ambady, 2010) an impression about her in the first few seconds.

The enormous power of deep learning is allowing us to predict outcomes at levels not previously thought possible, with input data that we used to consider too messy. But the leaders in deep learning are not in I/O or broader psychology fields, they are in engineering, computer science, physics, and the like. Psychologists must learn to use the power of AI-based solutions.

Psychometric Foundations. The new era of deep learning necessitates a wholesale re-envisioning of psychological assessment. The foundation of psychometrics rests on Classical Test Theory, which was designed to measure psychological attributes, or “constructs” as psychologists typically refer to them. These constructs cannot be directly measured, and so their existence must be inferred (Crocker & Algina, 1986). The most basic formulation of test theory is that a construct’s true score is equal to the observed score plus measurement error.

Infer we have. Psychologists have had a veritable field day creating every construct you can imagine, from the well-known (e.g., Extraversion) to the redundant (e.g., Grit). The problem with constructs is simple: they do not exist. Neuroscientists have not uncovered a group of brain cells that uniquely control Extraversion. The

construct of Extraversion is but a figment of a scientist's (in this case, Carl Jung's) imagination. A score on Extraversion may be useful as a broad descriptor of a subject's behavioral tendencies but modern psychological science can do better than simply averaging your score on five or so items that are nearly identical.

Topics to be discussed in this paper include: Test Theory Limitations: Test theory inherently promotes the idea that a construct has an actual "true score", as if it is an actual discoverable thing; Self-Report vs. Behavioral Sample: Moving towards scorable behavioral samples rather than self-report inventories; Fleeting Standardization: AI and related Big Data techniques use any and all available data, even if standardized indices are not available for each individual. We must develop techniques to handle unstandardized sets of data.

Starting with Data. It is naïve to believe that the human brain can reliably identify causal pathways among inordinately complex variable sets. Should the maximum complexity of our theoretical abilities set the upper limit on what we allow ourselves to study? Or should we study reality as it is, using all available means, and do our best to rationally explain what we observe? The path of strong science is the latter, not former. We cannot subjugate our newfound explanatory capability to the frailty of our minds. The paper will present several examples of stable predictive relationships that are not, at first blush, intuitive.

Moving Forward. With big data and deep learning, we now have a direct power to strongly predict real world outcomes that far surpasses the capabilities of historical methods, and reality is often vastly more complex than our simple, often linear and stereotype-based human theories can model.

Psychologists must embrace these new capabilities and rise to the challenge that they represent, and in so doing build a new paradigm of psychometrics for the deep learning era.

*Eric Sydell (Shaker International)*

## Independence E

### Social Media and Well-Being

#### PAPER SESSION

##### News Consumption on Social Media: Uninformed but Unaware

Despite the prevalence of social media in the media landscape, we still fail to understand how social media affects news consumption. Individuals are challenged by the information overload online (Gazzaley & Rosen, 2016). They focus less on detailed news reports and read news snippets instead (Petty & Cacioppo, 1986). In this research, we analyze whether public relevance of content strengthens this feeling of being well-informed without reading. As social media users have a strong desire to stay informed about what others are doing (Przybylski et al., 2013), we expect that users want to feel informed about publicly relevant content. However, if publicly relevant content is not personally relevant, individuals rely on news snippets instead of reading detailed news reports.

Social media users infer public relevance from social signals, such as the number of likes (Pentina & Tarafdar, 2014). In order to soothe their fear of missing out, we expect them to feel better informed about publicly relevant content even if they have not read thoroughly. It follows that the effect of public relevance on perceived

attention is only partially mediated by objective attention (Figure 1). The size of the gap between objective and perceived attention is defined by the strength of the direct effect of public relevance on perceived attention. We expect personal relevance of content to close this gap: If content is personally relevant, individuals are motivated to read and objective attention increases (Petty & Cacioppo, 1990). Moreover, perceived attention is less influenced by public relevance and content surrogates (Petty & Cacioppo, 1986).

Results from a first online experiment on MTurk (N=669) show that public relevance strengthens perceived attention (0.204,  $p < 0.05$ ). Individuals want to feel informed about publicly relevant content, perceived attention increases. Surprisingly, the effect of public relevance on objective attention is negative (-0.308,  $p < 0.05$ ). When content is publicly relevant, individuals read less thoroughly. As objective attention further affects perceived attention (0.11,  $p < 0.001$ ), the overall effect of public relevance on perceived attention depends on personal relevance. Personal relevance not only has a positive effect on objective attention (0.308,  $p < 0.05$ ) but also weakens the effect of public relevance on perceived attention (-0.406,  $p < 0.01$ ). The gap between perceived and objective attention closes with increasing personal relevance.

To shed light on the negative effect of public relevance on objective attention, we extend the theoretical framework. We expect that if content is publicly relevant users rely more on content surrogates and substitute a thorough read of content. Drawing from social influence theory (Deutsch & Gerard, 1955), users are either under informational or normative social influence: Informational influence occurs if individuals observe others' behavior to decrease uncertainty in ambiguous situations (Bikhchandani, Hirshleifer, & Welch, 1998). Users perceive public relevance as a quality signal and adjust their expectations about content informativeness accordingly. They do not have to read to ensure content quality and rely on content surrogates. Under normative influence, users conform to the perceived norm that they infer from others' behavior (Latane & Darley, 1968). In order to conform, individuals simply adjust their beliefs about the informativeness of content surrogates. Whereas under informational influence effects of public relevance are mediated by expected content informativeness, under normative social influence these effects are mediated by perceived informativeness of content surrogates.

Results from a second MTurk experiment (N=801) demonstrate that public relevance strengthens perceived informativeness of content surrogates (0.288,  $p < 0.05$ ) while not affecting expected content informativeness (0.190,  $p > 0.05$ ). Users adjust their beliefs about the informativeness of surrogates without updating their expectation about the actual content. We further see that perceived informativeness of surrogates weakens objective attention (-0.202,  $p < 0.05$ ) and strengthens perceived attention (0.131,  $p < 0.02$ ). These results indicate that users are indeed under a normative influence to feel informed about publicly relevant content. In order to conform they adjust their beliefs about the informativeness of content surrogates and feel better informed by headlines and abstracts. We further test two boundary conditions and find that the gap between objective and perceived attention persists even when individuals are not confronted with the decision to like. However, the gap vanishes for active social media users who frequently react to content on social media and are aware that likes are not representative of reading (Table 1).

This research adds not only to social psychology research on digital news consumption and the impact of social media on well-being. It further extends marketing literature on electronic word of mouth and content sharing. By now, research analyzed engagement

on social media without considering effects on actual information processing. This research shows that social signals lead individuals to feel better informed without actually reading. The results are of interest for news providers and social media companies. By now, recommendations mostly focused on attracting attention on social media. To increase the number of visitors to websites however, posts should not only trigger liking but also reading.

*Andrea Bublitz (University of Zurich)*

### **Linking as Collective Thinking? Participant Content Linking as a Means of Organizing Knowledge Meaningfully in Online Discussions**

Online multi-participant communication platforms promote generation of collective knowledge, through the sharing of content such as information, thoughts and ideas by participants (Haythornthwaite et al., 2018; Kent, Laslo, & Rafaeli, 2016; Rice, Heinz, & van Zoonen, 2019). But participants can contribute more than just content: they can, and they often do, provide means for organizing the shared content through the linking of content units. The work presented here turns the spotlight to this process and its effect on the structure of collective knowledge.

Knowledge is created through organization of information in meaningful ways (Shuell, 1986). Classic models for organization of shared knowledge applied hierarchical taxonomies or cataloging systems as a means of introducing shared order, which requires top-down organizing activities to maintain unified classifications (Hjørland, 2003). However, in an era of distributed, non-hierarchical organizational formations for learning, sharing knowledge and collaborating, bottom-up knowledge organization is becoming an essential part of the collaborative process. Innovation and collaborative creation of shared meanings require an environment that is open, enabling, and inclusive, yet still organized. In the absence of effective mechanisms for self-organization, collaborative efforts can result in knowledge that is fragmented and incoherent (Scardamalia et al., 2017).

Online multi-participant communication environments differ in the topology they allow: threaded discussions, prominent in many topical forums, learning management systems, and some social media platforms, are characterized by hierarchical, fully or semi-chronological structures (Biasutti, 2017; Zhang, Verou, & Karger, 2017). This format imposes a rigid, progressively branching out tree-shaped structure, hindering the discussion's potential to converge or self-organize in a meaningful way (Hewitt, 2001). In contrast, other formats of online communication allow networked topologies to emerge, by letting participants connect posts or other units of information that they see as related. This action, that can assume different forms, such as hypertext, tagging, cross-referencing and direct linking, allows participants to introduce meaningful connections that may deviate from the sequential hierarchy, break the tree-shaped structure and affect the discussion's overall network structure. The continuous reshaping of the network enables a dynamic, bottom-up organization of the collaboratively produced knowledge.

Our research examines the organization of shared knowledge through online multi-participant networked discussion. Inspired by current views suggesting that the organization of individual semantic memory is characterized by small-world structures that facilitate navigation while preserving complexity (Abad et al., 2015; Baronchelli et al., 2013; De Deyne et al., 2016; De Zubicaray et al.,

2011), we searched for similar structures in collaboratively formed knowledge to better understand the role of linking. Importantly, the emergence of small-world and any other complex structures requires non-hierarchical linking ("cross-linking").

Procedure and findings: We analyzed data from seven unmoderated discussions, held as part of blended academic courses (8-10 weeks, 24-82 participants each) and conducted in a platform that enables participants to "cross-link" pre-existing posts.

Graphs for network analysis were generated from discussion log-files with R igraph package (Csardi & Nepusz, 2006). The graphs were tested for small-worldness using the *S* measure, which compares the proportion between a graph's clustering coefficient and average path length to that of a random graph of similar size and order (Humphries & Gurney, 2008). Of the seven networks, six met the *S* criterion for small-worldness. In these graphs, the cross-links also reduced the number of modules, suggesting that cross-linking had been applied to connect posts that were further apart in the graph, perhaps representing the joining of ideas.

Next, partial networks representing the progression of each discussion in weekly intervals were extracted. Their analyses revealed a consistent rise of the small-world index, indicating that the participants' organizing activities were gradually molding the networks towards small-worldness.

Conclusion and implications: The findings suggest that cross-linking in multi-participant online discussions can assist in organizing knowledge, through bottom-up shaping of the discussion network structures.

With the discussions' progression and adding of more cross-links, most of the graphs advanced towards a small-world structure. Still, the findings imply that this is not an inherent feature of the discussion: not all of the discussions examined exhibited the small-world quality. As small-world structures have been suggested as optimal for learning and integrating new knowledge (Rodi et al., 2015), these measures may be indicators of the quality of group-level knowledge organization processes.

By viewing the cross-links as content organizing mechanisms, our work emphasizes the potential of participants as creators and organizers of collective knowledge, or, in other words, of linking as collaborating thinking.

*Noa Sher (University of Haifa)*

### **Digital Perceptions of Inner Peace: A Content Analysis of Meditation Depictions on Instagram**

This study provides evidence as to how meditation is represented on one of the most popular social media platforms currently in use by Americans: Instagram. More than one out of every four adults in the U.S use Instagram (Greenwood et al., 2016) to explore images produced by the platform's 500 million active daily users. As an interactive hub of consumers, brands, and institutions, the platform serves as a dynamic information space where para-social relationships occur, identities are cultivated, and perceptions of reality are informed. A number of studies have examined that platform's effects on psychological wellbeing (Hampton et al., 2015; Royal Society for Public Health, 2017; Tiggeman et al, 2018; Krutzen & Won, 2019). Perceptions of specific health behaviors in relation to Instagram depictions have also been explored (Tiggeman & Zaccardo, 2016; Phua et al, 2018), but none yet have investigated the growing corpus of contemplative practice content that can be found there.

While estimates of the percentage of Americans with a meditation practice differ, the American Physiological Association (2017) reported a rise in popularity to 12% across both meditation and yoga. Similarly, the Centers for Disease Control (Clark et al., 2018) reported 14% of American adults are now meditation practitioners, up from 9.5% in 2012. While increased interest in this realm of wellness is evidenced by Google search trends and content volume on social media, actual behavior change is further underscored by the market's economic growth as the global wellness industry grew 12.8% to \$4.2 trillion between 2015-2017 (Global Wellness Institute, 2018). As part of this wellness boom, there's been a surge in technology firms such as Headspace and Calm which aim to make meditation more accessible through mobile apps with guided instruction.

In the US, this boom may be stirred in part by the growing mental health crisis as the increase in those affected by chronic stress presents a pressing public health concern. The APA (2017) found an increasing number of Americans report feeling the effects of stress, but that stress and stress symptoms are unevenly experienced by different genders and races. What's more, a 2018 survey of 6,700 American adults found 57% report being paralyzed by stress, yet only one-third of those who acknowledge stress as a challenge report also visiting a doctor for stress-related symptoms (Everyday Health). Multiple factors have been identified for this troubling gap between experience and treatment, with substantial research pointing to the public's widespread perpetuation of mental health stigmas (Link & Phelan, 2001) as a major barrier to help-seeking behaviors (Leong & Zachar, 1999), in addition to treatment accessibility challenges consequent of time constraints, physical health, and financial limitations (Cohens Veterans Network et al, 2018). At the same time, a robust and growing body of literature establishes the effectiveness of meditation in managing stress (Oman et al., 208; Redstone, 2015; Munoz et al, 2018).

In light of such widespread mental health challenges, but improved accessibility to meditation as a therapeutic intervention, it's curious why adoption still remains relatively low as compared to other coping mechanisms. One avenue for understanding this reality is to examine perceptions propagated by online meditation content as communications scholarship has established clear links between the media one is exposed to and their perceptions of reality through first- and second-order effects of cultivation theory (Gerbner & Gross, 1976). This theoretical framework explores how cumulative exposure to specific media depictions subsequently inform ones' real-world beliefs on that topic. As an initial step toward further investigations using this model, this study delivers a baseline of digital depictions of meditation.

Specifically, this study examines Instagram content developed by public actors. In accordance with the new 2018 Instagram application programming interface (API), the full corpus of Instagram posts tagged with the term "meditation" made between February 18, 2019 and February 24, 2019 was collected using a third-party platform. Of the resulting 60,105 posts collected over this one-week period, a sample of 400 posts were randomly selected to support subsequent analysis at 95% confidence level. A manual content analysis is being conducted of both visual components (e.g. gender, race, and attire of human subjects, text on screen, and environment) and posting account to identify the percentage of brands posting as compared to individuals, as well as the gender of individual posters.

This on-going content analysis is not only developing a critical understanding of how meditation is portrayed online in regard to

gender, race, and environmental accessibility, but also how such portrayals are related to online behaviors of engagement such as "likes" and "comments." What's more, findings may then be used for more intensive examination of cultivation effects regarding meditation by understanding how excessive Instagram users' perceptions differ from those of their more moderate usage counterparts.

*Sarah Parker Ward (Emerging Media Studies, Boston University)*

### **Confirmatory Factor Analysis to Analyze Instagram Questionnaire Factor Structure Fit Across Cultural Settings**

**Introduction:** Approximately 65% of American adults identify as regular users of social networking sites (SNSs; Perrin, 2015). While certain demographics may be more likely to use SNSs, research has shown that the use of SNSs has grown significantly across all demographics over the last decade (Kaplan & Haenlin, 2010; Perrin, 2015). This rise in SNS use has coincided, and often been attributed to, the rise in prevalence and accessibility to smart phones (Suki, 2013). Instagram in particular is a popular SNS with over 300 million registered user accounts in 2015 (Ibba, Orrù, Pani, & Porru, 2015). Instagram has been documented as serving multiple different purposes, marketing and branding (Colliander & Marder, 2018), self-promotion and building self-reputation (Alshawaf & Wen, 2015), creating connections with other individuals (Ibba, et al., 2015), and much more. One thing that makes Instagram particularly noteworthy, though, is that it is unlike any other SNS in the sense that it relies primarily on pictures to communicate as opposed to predominantly text-based communications. As a result, Instagram has been noted as providing a raw form of visual culture for its users (Bradley, 2015).

**Purpose:** The Instagram and Well-Being questionnaire (Wong, 2017) is a non-validated, 19-item self-report inventory designed to assess Instagram and social media use. Item 14 of the measure contains 40 sub-items specifically assessing reasons for an individual's Instagram usage. Each sub-item contains a potential reason-for-use and participants respond on a Likert-scale based on how much they use Instagram for that reason. In prior research, an exploratory factor analysis (EFA) was conducted using a sample of 367 undergraduate students at the University of Sydney (Australia) and results demonstrated a three-factor structure: (1) sense of belonging, (2) self-expression, and (3) documentation/curation (Wong, 2017). The purpose of this study was to assess the goodness of fit of this factor structure in a sample of undergraduate students from the United States, oversampling African Americans.

**Sample:** Data were collected from 735 students at two public universities in the Southeastern region of Virginia, USA. One of the recruitment sites is a larger research-based institution with a fairly diverse student body and the other is a mid-sized liberal arts Historically Black university. After cleaning the data and removing cases due to significant missingness, 520 participants were retained for the analyses. The resulting sample self-identified as female (81.7%), male (18.1%), other gender (< 1%), African American (42.1%), Caucasian (37.9%), Alaskan Native (< 1%), Native American (< 1%), Asian/Asian-American (4.2%), Latino/a (7.1%), Hawaiian Native or Pacific Islander (1%), and multiracial (7.1%). The average age of the sample was 22.51 years old. All participants reported current use of Instagram.

**Method:** Confirmatory factor analysis (CFA) was conducted using Mplus to assess goodness of fit for the previously established three-factor model. The means and variance adjusted weighted least

squares (WLSMV) analysis method was used for this study, as this is the optimal method for categorical and Likert-scaled data (Beauducel & Herzberg, 2006).

Results: Goodness of model fit was assessed:  $\chi^2(87) = 1542.82$ ,  $p < .001$ ; RMSEA = 0.19, 95% CI [0.18-0.20]; CFI = 0.77; TLI = 0.73. Follow-up exploratory analyses were conducted to determine if the three-factor model maintains fit when the sample was partitioned into African American students only and Caucasian students only. In both cases, though, the three-factor model did not meet statistical criteria for exact or approximate fit.

Discussion: According to the results of this CFA, the three-factor model previously established (Wong, 2017) does not meet criteria for exact or approximate model fit for this data. While this could be due to a multitude of reasons, poor model fit could be due to the cross-cultural component of the study. Because the original three-factor structure model was found within Australian participants, it is possible that the same model may not maintain reliability and fit across different cultural samples. Given that the three-factor model continued to demonstrate lack of fit after partitioning the sample by race, it is most likely that global cross-cultural differences are influencing the differences in model fit.

*Daniel Schaffer (Virginia Consortium Program for Clinical Psychology)*

## Independence FG



### Measuring Psychological Variables using Mobile Sensing Technologies: Modeling Big Data and Implications for Research and Designing Intelligent Support for Well-Being and Productivity at Work

#### SYMPOSIUM

The world of work is changing. With IT and educational advances, work productivity has increased over the past century. On the other hand, the continuous connection to our technological devices and the high demands for unrelenting, persistent communications (e.g., responding to emails/instant messengers, doing so promptly) has added to perceived workplace stress. Poor workplace wellbeing has been a serious problem in individual employees, institutions and the nation. Beyond its detrimental impact on employee physical and psychological health, stress at work can lead to reduced job performance, increased absenteeism, presenteeism, and even counterproductivity. However, measuring wellbeing and performance objectively and quantifying changes with time or job demands is challenging. Both self-reports of stress and wellbeing are fundamentally subjective. Moreover, these measures are usually acquired at discrete time points in non-ecological settings such as annual workplace or stress evaluation surveys. Technical innovations in machine learning, and wearable sensors provide critical opportunities to advance measuring workplace wellbeing and linking them to performance trajectories and designing intelligent support for employers and employees.

This symposium presents research and findings from the mPerf project: a multimillion-dollar effort for modeling employee wellbeing and everyday work behaviors using unobtrusive, wearable mobile sensors and modeling the resulting data using innovative Big Data modeling approaches. Psychologists, computer scientists, engineers

and data scientists collaborated on building, deploying and utilizing a variety of wearable mobile sensors to gather psychological data from several hundred employees from several organizations. Data collection lasted for 10 weeks. This symposium will provide an overview of the sensing technologies used, the data analytic approaches utilized, and some early findings from the project. Practical learnings for other researchers and practitioners wishing to use sensors to measure psychological variables and/or to model voluminous, time-intensive data from devices will be shared.

This symposium will explore the following broad themes through the oral presentations and a joint panel discussion.

1. Using multimodal, temporally intensive mobile sensor data to measure and assess employee wellbeing: which mobile sensor markers are useful as indicators of work relevant psychological constructs? Which mobile sensor devices yield useful, predictive data?

2. Ensuring the reliability, validity, replicability, and generalizability of machine learning algorithms to model wellbeing and work performance of employees: How can researchers apply machine learning models to obtain data-driven indicators of work-relevant psychological constructs? How can these indicators meet standards for reliability, validity, replicability, and generalizability to be useful in applied settings?

3. Building reliable and accountable decision support algorithms to ameliorate employee stress, improve wellbeing and worker productivity: What technological innovations are necessary to build interventions using input data from wearable mobile sensors? How do we evaluate their safety and accuracy?

The first presentation (Ones, Dilchert, Stanek et al.) will provide an overview of the mPerf project, its aims, sensors, and approaches. The second presentation (Wiernik, Wang et al.) will highlight the psychological variables measured and the theoretical rationale for selecting various, specific sensor based indicators for assessing them. The third presentation (Sano et al.,) will present results from modeling of mobile sensor data for assessing intrapersonal variation in employee wellbeing and work performance constructs. The fourth presentation (Ones & Sano et al.) will describe how mobile sensor data can be used to design intelligent support systems for wellbeing and productivity at work. Across these presentations, the symposium attendees will gain insights into the types of mobile sensors that can be unobtrusively utilized with working populations to gather Big Data for measuring psychological wellbeing and work behaviors. They will also gain an understanding of opportunities and challenges such data hold for reliably and validly measuring psychological constructs of interest. Data modeling approaches and choices utilized will highlight innovative solutions to challenges encountered. Replicability and generalizability of the models generated will be specifically discussed. The presentations in this session are aimed at engaging a multidisciplinary audience with the question of how mobile, wearable sensor data can be used to assess and support wellbeing and productivity in work settings. The future of work relevant feedback, behavior modification, and individually targeted wellbeing and productivity interventions include mobile sensor technologies.

The proposed symposium seeks to bring together an interdisciplinary group of researchers and practitioners from academia and industry to discuss the unique opportunities and challenges for developing effective, ethical and trustworthy sensing, algorithms and applications for future of work: measurement, prediction and intervention of employee wellbeing and job performance. This symposium hopes to strengthen the bridge between ubiquitous and affective computing and psychology research communities, to present the

latest applications of cutting-edge computing tools to psychology studies for designing future of work and to consider opportunities to expand the use of these tools in workplace settings.

#### **Deploying Unobtrusive, Persistent Mobile Sensors to Measure Psychological Constructs**

*Deniz Ones (University of Minnesota-Twin Cities)*

#### **Mobile Sensor Markers: Theoretically Oriented? Data-driven? Replicable?**

*Yilei Wang (University of Minnesota-Twin Cities)*

#### **Measuring and Predicting Psychological Well-Being at Work Using Mobile Sensors and Machine Learning**

*Akane Sano (Rice University)*

#### **Potential of Mobile Sensor Data and Machine Learning for Well-Being and Productivity Improvement: Monitoring, Feedback, and Interventions**

*Deniz Ones (University of Minnesota-Twin Cities)*

---

### Farragut Square

## Learning and Education in the Digital Age

### PAPER SESSION

#### **Technology Trends:**

#### **A View of Evolution for Educational Technology from 2011 to 2017**

Educational technologies have captured the attention of researchers, policy makers, and parents. Each year, considerable effort and money are invested into new technologies, hoping to find the next effective learning tool. However, technology changes rapidly and little attention is paid to the changes after they occur. Greater trends should be aware as proper awareness not only allows policy makers and the public to spend their educational capital more wisely by avoiding the adopting of devices that are unlikely to persist but also allows researchers to better understand those technologies that are likely to be used in a broad range of classrooms rather than those adopted by the few techno-enthusiasts. The Horizon Reports (HRs) are a global ongoing research report exploring technology trends that are likely to have an impact on formal education. Each year, Horizon Report finalizes six technologies that predicted to influence K-12 teaching and learning in the following one to five years across three periods: near-term (the year of the report), mid-term (2-3 years), and far-term (4-5 years). So far, only Martin and colleagues' study has analyzed the Horizon Reports. Marin et al., (2011) provided an evaluation of the most important technology trends in K-12 education across 2004 to 2010 by comparing the technology adoption rates predicted by the HRs with published articles in Google Scholar using bibliometrics analysis. However, since Martin et al., (2011) work, no recent study has been done on reporting and examining the changes in educational technology after 2010. Further, Marten et al.'s work somewhat ignored an assumption underlying the HRs about the connection between society and education, that broader societal trends in research and practice determine the educational community's mainstream technology usage.

To address this gap, this paper adopted Martin et al.'s (2011) methodology but with the latest HRs from 2011 to 2017. The methodology involved the following stages:

1. Seven Horizon Reports were gathered from 2011 to 2017 and technologies predicted in each report were recorded.

2. Based on the records, a visual representation of the HRs' predictions was made. These visualizations provide a clear picture of all the technologies predicted in the past seven years.

3. Similar technologies across all the reports were grouped into clusters and visual representations (same in stage 2) were created for each cluster. These clusters are used in the subsequent bibliometrics analysis.

4. Using the newly created clusters, the evolution of educational technologies across 2011-2017 were analyzed and discussed for each group.

5. The fifth stage involved the bibliometrics analysis. Bibliometrics analysis was used since the number of publications on a given educational technology are an index of its importance (Norton, 2001; Morries, 2002). Google Scholar was used as the bibliographic database as it provides metadata of scholarly literature across disciplines and connects repositories worldwide. The search process involved keywords (based on the clusters identified in stage 3), year of publication, and narrowed down only if "education" or "learning" appears in the title. Since total number of education related publications available in Google Scholar changes every year.

6. To assess the accuracy of the HRs predictions, the trends predicted by HRs were compared to their impact from stage 5.

The results suggest that mobile technologies were the most impactful in the HRs from 2002 to 2010. Maker technology, analytics technology and games were predicted to impact education in 2015-2018, 2014-2019 and 2012-2016, respectively. Simulation technologies' impact was predicted to increase and would continue to influence learning along with other promising emerging technologies like AI. Thus, the HRs predictions continue to highlight pervasive technology (mobile) while recognizing the emerging technologies (VR, AI, Maker). The bibliometrics revealed that the HRs' predictions were generally correct: there was a trend towards games in the early years, maker and simulation technologies in the middle years after 2015, mobile and analytics technologies trended consistently across the period, and emerging technologies (e.g., AI) are tending to influence learning in the future.

What can be noted is how both the year of prediction and the publication rates for emerging technologies seem to coincide with availability of the technology at a consumer level (i.e., affordable). Consumer level maker and simulation (VR, AR) technologies became available the same year they were included in the HRs and their publications rates increased in the two years following their commercial availability. This suggests that trends in educational technologies are driven more by their availability than their educational affordances.

This work provides a picture of educational technology trends in the past and near future. The results not only reflect the evolution of some technology clusters, but can also serve as a guideline for researchers, policy makers, educators and parents to make decisions on technology adoption in the future.

*Sabrina Shajeen Alam (McGill University)*

### Detecting Hidden Patterns of Engagement in a Digital Cognitive Skills Training Game: Implications for User-Centered Design

Sustained cognitive engagement is thought to be critical for the effectiveness of computerized cognitive skills training interventions (Anguera & Gazzaley, 2015). We sought to identify patterns of engagement over the course of a game-based cognitive skills training intervention study and to determine whether the unique trends in game-play performance would emerge that could not be explained by other factors. As a data-driven approach to understanding human-computer interaction within the context of a cognitive skills training game, we hoped the findings would (1) identify systematic variation in patterns of engagement with the game, and (2) inform future iterations of such games so that they may become more user-centered.

**Methods:** Participants (N=144) were from middle and high schools near a metropolitan region in the Northeastern U.S. (Mage=13.7 years, SDage=1.1, %female=45.5, 41.1% reported speaking English at-home, 56.0% reported speaking Spanish at-home). All participants were enrolled in the study in the fall 2018 and spring 2019 during which time they were expected to complete six 20-minute game-play sessions over the course of four weeks. In this analysis, we examined a subset of the intervention data from a counterbalanced treatment condition wherein participants played a custom-designed game created to enhance the executive function skill of inhibitory control. The game has a similar basic design as a GO/No-GO task. Participants played a game designed to target inhibitory control skills for half of the session (10-minutes). In one condition, participants played the inhibitory control game first during each session, followed by another game for the same duration of time, while the other condition the order of the games were reversed.

Computer log data was recorded during game-play, and included information about the target type (e.g., target or distractor), the response, and the reaction time in response to the target. This particular game involved the presentation of both targets, distractors, as well as delayed targets and distractors. For the purposes of this analysis, delayed targets and delayed distractors were excluded from the analysis. As a measure of performance, we used the sensitivity index, *d*-prime, where a greater value indicates a more precise detection of the targets. We examined participants' *d*-prime over the first four of six sessions. We chose to examine the first four sessions both for the purpose of building a model that could be sensitive enough to find differences in patterns of performance prior to the cessation of the intervention, and for the practical reason that not all participants completed six sessions.

**Results:** We were interested in determining whether there may be differences in participants' patterns of performance. Latent class mixture-modeling with a linear link function was used to examine patterns in performance using *d*-prime as an outcome. Post-fit indices were examined to assess the goodness-of-fit of models with one to five cluster solutions. The *lcmm* package in R was used to conduct the analyses (Proust-Lima et al., 2017).

Comparison of the goodness-of-fit of the models indicated that clustering with 2-groupings provided optimal fit. Fit indices were improved in the 2-group model as evidenced by lower AIC, BIC, and maximum loglikelihood values, see Table 1. The majority of participants fell within the same cluster (nk1=118), which was characterized by a gradually increasing in *d*-prime over time. By contrast, the average trajectory of a subset of participants (nk2=25) was characterized by low initial scores, followed by a progressive improvement in

*d*-prime over time, see Figure 1. Results of a Chi-square of independence confirmed that the clusters were not related to the counterbalanced condition or school-level, Chi-square(3, 143)=5.86, *p*-value=.12.

**Discussion:** Past research on the use of computerized cognitive skills training interventions has yielded inconsistent findings regarding the efficacy of such trainings (Jacob & Parkinson, 2015; Simon et al., 2016). While it is recognized that the extent to which trainees are maintain active engaged over the course of the training is a likely source of variation in the efficacy of such trainings (Gathercole et al., 2012), few studies have attempted to use a data-driven approaches to studying engagement within the context of these trainings. This exploratory study aimed to address this need by shedding light on patterns of performance that may otherwise be undetected by the naked eye.

Further analyses may seek to identify whether these latent classes are meaningfully associated with improvements in measures of the the target cognitive skill of inhibitory control. In addition, better and ongoing metrics of performance may be developed that can formatively assess student progress within the context of the game (Shute & Kim, 2014) and could potentially be used to adaptively adjust the level of difficulty of the game (Holmes et al., 2009).

*Teresa Ober (The Graduate Center CUNY)*

### The Art of Looking in the 21st Century

Camera-enabled smartphones have transformed how individuals engage with their visual surroundings. Gómez Cruz and Meyer (2012) have asserted that smartphone photography represents a "reconfiguration of the definition of what photography means in the digital era," made possible by the fusion of camera, connectivity, Internet browsing, and geolocation. However, very little anthropological research has focused on smartphone photography in distinct visual contexts. (A1) In this paper, I investigate how and why individuals choose to mediate their visual world with smartphone photography, using the museum context as a paradigm for thinking about digitally mediated visual engagement more broadly. I suggest that key cognitive science insights can help those approaching social scientific studies of technology and exemplify this suggestion by using cognitive science literature to inform my ethnographic account of smartphone photography in museums.

Perhaps because of their inherently visual nature, museum environments serve as a platform for academics, writers, and other thought leaders to engage with the authenticity of photographically mediated experience more generally (Rosenbloom, 2014; Gilbert, 2016; Kozinets, Gretzel, & Dinhopf, 2017). Museums have also been featured in the press for capitalizing on smartphone photography trends, mounting shows and changing the types of art presented so as to attract photographing visitors (Gamerman, 2017). The frequent discussions regarding smartphone photography in museums—highly visual and experiential contexts—make museums an interesting domain for exploring broader questions of why and how individuals choose to smartphone photograph.

Scholars have attempted to address the motivations for photographing since well before the dawn of the smartphone era (Sontag, 1978; Barthes, 1980). Several themes have emerged out of this research that, when fused with foundational cognitive science insights and the affordances of smartphone technology, can inform anthropological study of smartphone use in museums. While I anticipated that a variety of themes would emerge from ethnographic

study due to the different personalities of those studied, I also predicted that several themes would appear to unify individuals in their photographing behavior at the museum: 1) the desire to remember (Routledge, Arndt, Sedikides, & Wildschut, 2008; Schacter, Addis, & Szpunar, 2017), 2) the rewarding nature of social engagement (Tamir and Mitchell, 2012), and 3) habitual connection to smartphones and their platforms, perhaps guided by the social engagement and memory capabilities they provide (Sussman, Lisha, & Griffiths, 2011; McClure, York, & Montague, 2004).

(A2) To explore how memory, social engagement, and habituation motivate smartphone photography practices, I used covert observation, semi-structured interviews, and digital ethnography of photo-sharing behavior on Instagram to investigate smartphone photography in four distinct museum contexts: art (The Ashmolean Museum), natural history (The Oxford University Museum of Natural History), anthropology (The Pitt Rivers Museum), and cultural history (Jewish Museum Berlin). I devoted an average of ten hours to covert participant observation in each museum before conducting 36 semi-structured interviews; I selected interview participants who mediated their visual experience with smartphone photography and those who did not, attempting to control for age and gender. (A3) I chose to analyze the interview transcripts via an informal form of attribute coding (Saldaña, 2009).

(A4) I conclude that a desire to remember, the rewarding nature of social engagement, and the habitual connection to smartphones and their platforms motivate smartphone photography in museums. (A6) Specifically, those who believed that smartphone photography precludes authentic visual experience opted for unmediated viewing, while those museumgoers who sought to preserve their visual environment for future recall turned to smartphone photography to do so. Additionally, the majority of those who used smartphones to photograph in museums did so out of the habitual desire to photograph in response to their sensory environment, or to share with others via social media or SMS. An unexpected theme also emerged via observations at the Jewish Museum Berlin and of its Instagram feed: emotionality can influence manifestations of smartphone photography in the museum context.

The camera smartphone should not be considered an iterative phase of photographic development, but rather as a distinct technology that affords novel photographing and visual engagement behaviors. (A5) This paper not only explores how the smartphone camera can affect how people engage with historical, artistic, and cultural artifacts and narratives, but it also begins to unpack how smartphone photography has changed conceptualizations of authentic human behavior and experience via mediated engagement with the visual world.

*Claire Leibowicz (The Partnership on AI/Oxford Internet Institute)*

#### **Using Technology to Self-Monitor — Demonstration of the I-Connect Mobile Application and Student Outcomes**

I-Connect is a data driven self-monitoring phone application. It takes traditional pen and paper self-monitoring to the next level by allowing the user to monitor their behavior digitally, independently and in real time! The current I-Connect user population consists of teens and young adults with ASD and/or ADHD, but I-Connect could be used to help any individual that struggles with staying on task in the classroom or during routine activities. The application allows users to monitor at school, work, home or during any activity. All monitoring

questions and the variable rate at which they are asked are customizable. Monitoring questions are broken down into three categories; on task, appropriateness, and comprehension. These categories are customizable, for example “are you on task?” is a preset question but in this category this question could be customized to say “are you completing your work?” I-Connect then ask the assigned self-monitoring questions at a variable rate that is an appropriate fit for each individual user’s needs. Some students may need to monitor every 5 minutes, while others might only need to be asked 2-3 times a class period. These custom self-monitoring locations, questions, and the variable rate at which they are asked are set up on the back end of the application by the educator or mentor and they aren’t accessible by the user. I-Connect uses the data collected by the responses to create individualized charts for each user. When assigned, these charts are accessible by mentors, teachers, parents and other stakeholders in the user’s life. These charts are useful for writing IEP’s and/or seeing how the student has improved behavior while using the application. The I-Connect application is a tool that can be used with one individual user or an entire school district! In this presentation attendees will get an overview of our experience with the I-Connect Application intervention and a summary of the outcomes it has created. We will also provide some history, context, and rationale for implementing self-monitoring. We will walk attendees through our research studies and show video examples of our experience. Finally, we will give attendees an opportunity to interact with and experience the I-Connect application by providing devices with pre-set user accounts.

*Kajsa Mullenix-Mohammed (The University of Kansas)*

#### **Independence HI**

#### **Teletherapy**

#### **PAPER SESSION**

#### **The Impact of Product Design on User Engagement with the Therapeutic Process in e-Mental Health Interventions**

Context: The use of self-guided mental-health apps has the potential to increase access to care in a scalable manner by reducing the costs associated with service uptake. However, the impact of digital interventions is limited by their ability to engage users in therapeutic activities and to support user adherence to the therapeutic process. There is a body of literature examining how intervention design may facilitate engagement and behavior change, however, the understanding of these relationships is still limited by our ability to examine user engagement in the real world, while comparing large numbers of products within the same study.

Objective: This presentation aims to examine benchmarks of real world use of mental health apps and whether the qualities of product design, research evidence, and publicly available data predict real-world user engagement with mobile and web-based self-guided e-mental health interventions.

Methods: First, utilizing a panel that presents an aggregated non-personal information on user engagement with websites and mobile apps across the world we examined benchmarks of user engagement with mental health apps (depression, anxiety, and emotional well-being). Second, we used the same panel to examine the



correlations between user engagement in the real-world, research evidence, and different quality aspects of product design—rated prior to empirical testing using the Enlight suite of quality scales (Usability, Visual Design, Content, Therapeutic Persuasiveness, and Therapeutic Alliance).

Results: Findings indicate that across a wide range of incorporated techniques and clinical focuses the Median 30-day user retention ranges between 0.0% and 8.9%, with a difference found favouring trackers and peer-support apps in comparison to breathing exercise apps ( $P_s < .001$ ). Among active users, daily minutes of use were significantly higher for mindfulness/meditation and peer support apps than for apps incorporating other techniques (tracker, breathing exercise, psychoeducation;  $Z_s \geq 2.11$ ,  $P_s < .05$ ). The median of app 30-day retention was 3.3%.

In terms of product design, most self-guided programs that had a high quality rating in terms of Content did not have high quality of engaging features or Therapeutic Persuasiveness (the incorporation of persuasive design and behaviour change techniques) within the digital product. Moreover, results point to moderate correlations between Therapeutic Persuasiveness, Therapeutic Alliance, and the three user engagement variables ( $P_s \leq .03$ ). Visual Design, User Engagement, and Content demonstrated similar degrees of correlation with mobile app engagement variables ( $P_s \leq .04$ ), but not with average usage time of web-based interventions. While several product quality ratings were positively correlated with research evidence, the latter was not significantly correlated with real-world user engagement.

Conclusions: Findings emphasize the difference between evidence based content and evidence informed design of digital products. Quality aspects of product design should be taken into account when discussing user engagement with intervention or clinical outcomes.

*Amit Baumel (University of Haifa)*

### **Ensō: Using A Web-Based Application to Foster Psychological Health In Health Care Providers**

Healthcare providers, including psychologists, experience occupational stress which has been shown to lead to secondary trauma, burnout, and associated psychological suffering, such as emotional exhaustion and depersonalization as well as poor outcomes and lower satisfaction in patients (Shanafelt et al., 2012; Peckham, 2015; Salyers et al., 2016). Ever-growing strains in healthcare, coupled with a competitive and demanding work environment, have led to numerous negative psychological consequences including increasing rates of burnout with 54% of physicians meeting criteria (Shanafelt et al., 2015) and the suicide rate of providers increasing (Schernhammer, 2005). There are strong links between a clinician's level of distress and the likelihood of displaying unprofessional behavior. Increased awareness has resulted in more frequent reports of distress, encouraging many healthcare organizations to respond by creating a "healthy work environment."

Although clinician burnout has received increasing attention, the impact of proactively screening and utilizing early psychological and contemplative interventions in healthcare environments remains comparatively underexplored (Penberthy, et al. 2018). It is not only important to identify those who suffer within the workplace, but also to ensure the distressed individuals are actively invited to be cared for and that prevention programs are available. The importance of

this point is made evident in a 2012 study revealing that 78.3% of the distressed physicians surveyed had not previously thought about seeking professional help for depression or burnout (Fridner et al., 2012).

The modern healthcare system may benefit from an increased focus on clinician well-being and how to achieve and sustain it within the stressful field of healthcare, as well as strategies to help transform the healthcare system to allow sustained engagement. Well-being should be considered as more than simply the absence of distress. Programs teaching mindfulness, effective communication skills, and stress reduction techniques may be key in helping to establish a resilience and effective group of healthcare workers. Ongoing screening, assessment, feedback and intervention are needed to sustain continued engagement and a healthy work environment. In addition to these interventions, longitudinal studies examining both clinician distress and well-being are needed in order to identify and implement the interventions that have been proven successful. Systemic changes in the health care system or environment must also continually be evaluated and revised for optimum efficiency and engagement of all providers and patients. Clinician engagement in a mindful communication program is associated with both short and long-term well-being and attitudes associated with patient-centered care (Krasner et al., 2009). Additionally, mindfulness-based prevention programs for clinicians have been demonstrated to reduce burnout levels (Goodman and Schorling, 2012).

In this project, we built upon our previous experiences of teaching clinician well-being and implemented a smart web-based application to assess individual and environmental stress using validated screening questionnaires and provide personalized feedback and support, especially to those at risk for burnout. Individual scores on standardized assessments were tracked over time, as well as aggregated scores for the larger unit in which the individual provider worked. Feedback about the scores were used to instruct providers about effective preventative strategies. Systemic issues in the work environment were also identified and addressed simultaneously at the institutional level. The program used to collect screening data and provide customized feedback is called Ensō, which means enlightenment. Ensō is the screening and assessment arm of a larger health system program called Be Wise. Be Wise, a wisdom based assessment and training program, is designed and implemented to promote professionalism, engagement, and ultimately, a healthy and safe work environment by increasing awareness of and skills in such things as stress management, emotional intelligence, compassion, resilience, and effective interpersonal communication.

All clinicians in the health system were invited to access Ensō, the web-based application, to assess and track their perceived stress. They were then provided resources tailored to their screening scores to foster insight, and promote effective communication skills and self-care. Individual scores were combined with other scores from the same unit, clinic or department in order to obtain an overall stress score for the larger unit. Scores were tracked over time and appropriate interventions were implemented depending on individual and overall unit or clinic scores. Those individuals not endorsing distress were given personalized and immediate feedback, reinforced for the adaptive strategies used, and provided preventative educational resources, and invited to join a mindfulness-based stress reduction program. All clinicians were screened with empirically supportive and validated psychometric questionnaires used in current practice (Bech, 2012). These metrics were analyzed in

reference to employee engagement and patient satisfaction to determine correlations between stress and clinical outcomes. Overall, clinicians demonstrated high enrollment, engagement, and resource use. Continuous follow-up revealed decreasing burnout and distress scores and increased employee engagement. This intervention demonstrates the feasibility of a web-based tool to systematically screen and utilize psychological and contemplative interventions to proactively improve coping and interpersonal skills of clinicians in a large healthcare system.

*Jennifer Kim Penberthy (Department of Psychiatry & Neurobehavioral Science)*

### **Telerehabilitation Care: A Novel Approach to Meeting Clinical Needs of Veterans with Disabilities**

Telehealth is a convenient and effective way for patients to gain access to, and for clinicians to provide, specialty clinical services that are typically not available outside of major medical centers. In recent years, the VA has implemented a nationwide system to enable training in telehealth and to ensure quality in telehealth service delivery, facilitating services from more than 900 VA locations and in over 50 specialty clinical care areas.

A recently established VA telehealth program is the Telerehabilitation Enterprise Wide Initiative (TREWI), designed to increase access to specialty rehabilitation care services for Veterans in rural locations. Prior studies have identified that telehealth treatments for physical conditions are effective at improving functioning (Cottrell et al., 2017), supporting telehealth as an appropriate mechanism for delivery of rehabilitation services. This program includes 4 "hub" sites (Minneapolis, Richmond, San Antonio, Seattle) as well as smaller "spoke" sites. This hub/spoke model facilitates consultation between larger medical centers and VA sites with fewer rehabilitation providers. This program offers consultation to rehabilitation providers on effective development and implementation of novel telehealth services for Veterans. Providers involved in the TREWI program include physiatrists (physical medicine & rehabilitation physicians), occupational therapists, physical therapists, speech pathologists, social workers, and rehabilitation psychologists/health psychologists. TREWI has facilitated provision of numerous specialty medical and mental health services for Veterans with broad ranging medical conditions, physical and mobility limitations, and cognitive impairments.

This presentation will discuss the different methods of service provision (telehealth to community clinic, telehealth to home) and specialty services that have been offered through this program and present Veteran feedback on these services versus traditional care delivery methods. Specialty psychology services currently being offered through this program include brief rehabilitation psychology interventions targeting sleep disturbance, non-pharmacological pain management, and adjustment to disability. Other programs include traumatic brain injury intake assessments, telehealth adaptation of neuropsychology services, and interdisciplinary amputation clinics. These programs extend services that are currently only available at select, large medical centers to patients living in more remote locations. These services offer a significant benefit in terms of reducing travel time and preparation required in order to attend appointments, particularly for patients with significant physical limitations. These type of barriers can be particularly impactful on adherence

and participation in rehabilitation services that frequently require repeated visits (Cowper-Ripley et al., 2019).

Data that is currently being collected will be presented on patient satisfaction as well as the range of services offered to Veterans by providers within different healthcare disciplines. This presentation will additionally discuss ways of adapting traditional telehealth technology for Veterans with different physical, cognitive, and sensory limitations (e.g., patients with low vision, patients with limited hand function). Discussion will also include ethical considerations for weighing risks versus benefits of adapting specialty mental health treatments for telehealth, particularly with patient populations at elevated risk of suicide/self-harm or acute medical events (e.g., stroke, heart attack).

The data gathered from this novel program will help to inform future medical care for patients inside and outside of the VA who could benefit from receiving specialty medical and mental health services but reside far away from a major medical center. The ability to expand the boundaries of the types of services that can be offered to patients is essential for ensuring that patients are able to receive high quality care without the burden and cost of travelling long distances to receive that care. The VA has created an infrastructure that can help guide the expansion of telehealth beyond the VA system by identifying challenges and key supports that are essential for implementing telehealth services in a way that is ethical, clinically appropriate, and effective.

*Larissa Del Piero (Veterans Administration, Puget Sound)*

### **The Divided Psychotherapy and the Internet: Some Suggested Integration via Hybrid Models.**

For many years, the world of psychotherapy has been divided into different parties. One classical distinction is between the "Techniques party", and the "Relationship party" Each of these parties view different components as central towards bringing about changes. The arrival of the internet into the psychotherapy world could strengthen this split but can also integrate and produce some hybrid models. The symposium will focus on such hybrid interventions where the internet allows such integration.

A) Integration between the dynamic and CBT parties: In this part we will focus on the type of relationships that can exist via online communication. Liat will distinguish between CBT and Dynamic interventions. In both cases she will elaborate on the therapeutic alliance and share her suggestions how to enhance the alliance online. Liat will also describe a unique type of a hybrid intervention where two therapists are working with the same patient. One of the therapist is present with the patient in the same room, while the other one is working online, and administer the CBT interventions from a distance. For example once there is a need for exposure, the distant therapist might challenge the avoidant behavior of the patient, while the therapist who is in the same room supports the patient via this process.

B) Integration between CBT applications and human interventions. We will share their 20 years of experience developing online therapeutic applications. Their talk will cover few internet based interventions. One of the interventions "Beating the Blues", which was developed for UK and proved itself in many successful follow up studies. However, in retrospect, We will share their belief that hybrid interventions which include both face to face and self help online program, is the best combination for a new era of psychotherapy.

C) Integration between various models of family and couples therapy, We will speak about couple interventions, using meeting via the internet. They will review issues of setting, seating arrangements of the couple and the therapists. Their focus will be on a unique hybrid interventions were they see the couple in a regular face to face session but also allowing the couple to initiate a session in real time from their home when the couple is in a crisis.

D) Integration between group supervision types: face to face and group supervisions. will discuss the emerging trend of online supervision. Specifically will describe how we do group supervision from USA to China and we will also describe online team supervision. After discussing the pros and cons of cybersupervision, We will describe an hybrid model of supervision where the same group of supervisee receive online and face to face supervision.

*Arnon Rolnick (Rolnick's Clinic)*

Independence BC



## Dietary Behavior and Technology

### PAPER SESSION

#### Employing the Virtual Reality Buffet to Measure and Understand Parent Feeding Behavior

One important predictor of adult obesity is eating behavior during childhood, a process in which parents play a significant role. Accordingly, it is crucial to understand the psychological processes that underpin the feeding choices parents make for their young children. Part of the difficulty surrounding the study of parent feeding behavior hinges on the methods available to conduct this research. Behavioral measures are the gold standard, however, the use of real food is costly and wasteful, is typically conducted in a sterile laboratory environment, and process measures (such as order and timing influence) are difficult to collect. To address these shortcomings, we developed the Virtual Reality (VR) Buffet tool wherein parents create a meal for their child using an immersive VR buffet simulation. This tool outputs precise information on food choices that parents make for their child, alongside information about the behavioral process through which these choices were made.

We have validated the VR buffet in a sample mothers and fathers of 3-7 year-old children. Participants completed a lab visit where they served a virtual lunch of pasta and apple juice from the VR buffet for their child, played Tetris as a distractor task, and served a lunch of real pasta and apple juice for their child (counterbalanced order). Parents then chose foods for their child from a larger virtual buffet and completed a questionnaire. Bland-Altman plots reveal that 96% of the data is within the limits of agreement, although there is bias such that parents tend to serve slightly larger amounts of real pasta as compared to virtual pasta. Parents who selected larger servings of pasta for their child in the virtual buffet also chose larger amounts of real pasta. In addition, anticipated patterns were evident in the data (i.e., parents served more food to older children and to boys). We are also currently collecting data about the role of olfaction in the VR buffet experience which will be available to present at the time of the conference.

We have employed the VR buffet in several substantive trials to assess the influence of informational interventions on parents' feeding behavior. For example, we have shown that providing parents with

family history-based risk information about their child's risk for obesity in the future is associated with choosing fewer calories for the child on the VR buffet. We have also shown that this risk information induces guilt among parents, but that choosing healthier foods from the VR buffet can reduce those feelings of guilt. In addition, we have used the buffet to examine the psychological factors that influence food choices that parents make for their boys versus their girls.

The VR buffet also allows us access to very fine-grained, continuous data relating to parents' food choice process. We have shown how risk information can lead to different food choice outcomes via differences in the micro-behaviors that make up the food-choice process. For example, parents who receive risk information about their child show less 'choice inertia', that is, they are less likely to keep choosing more servings of a food once they have started, a behavior linked with portion size selection. Another use of this fine-grained behavioral information is a study in which we tracked parents' walking behavior within the VR environment and linked elements of their path to psychological process. More tortuous walking paths when choosing food were linked to reduced guilt about child feeding.

In all, the VR buffet is a validated research tool that provides an opportunity to explore both process and outcome in parent food choice for their children and the psychological processes that accompany them. There are several future directions planned for applications and development with respect to the VR buffet that will continue to push forward research in diverse domains such as social psychology, cyberpsychology, and nutrition.

*Susan Persky (National Institutes of Health)*

#### Validation of a Virtual Reality Buffet Environment to Assess Food Selection Processes in Young Adults

Obesity is a pressing global health concern with enormous societal costs (Ng et al., 2014). The prevalence of obesity among U.S. adults was 39.8% in 2015-2016 (CDC, 2017). The onset of dysfunctional eating behavior and eating disorders occurs in adolescence and young adulthood (DSM 5), and weight gain is common during the transition to college, where young adults make more independent food- and diet-related decisions (Vadeboncoeur, Townsend, & Foster, 2015). Thus, there is a need to examine food- and eating-related processes within this developmental period and population, which can ultimately inform prevention and intervention efforts.

Virtual reality (VR) food buffets have been shown to be a promising tool for both basic and applied research on eating and food-related processes. Researchers can create VR environments with high degrees of immersion and naturalistic stimuli, enabling better generalizability of findings to the real world (RW) compared to controlled laboratory studies (Forman et al., 2018). VR buffets allow researchers to manipulate environmental factors (e.g., the presence of other people, food color and texture) and to use a variety of automatic measurement tools assessing the order, timing, and other food selection details (Marcum, Goldring, McBride, & Persky, 2018).

The overall goal of the present study was to examine the validity and utility of the VR food buffet in assessing food selection. Our study extends previous VR food buffet validations by allowing participants to select from a wide variety of choices in a VR food buffet (10 sides, 2 soups, 15 entrees, 6 fruit and desserts, 9 drinks) and a RW campus dining hall buffet.

We proposed the following research questions: (1) Convergent validity: Will participants' food selections in the VR buffet be related to their RW food buffet selection (in Kcal, grams, fat, carbohydrates,

protein)? (2) User experience: What were participants' perceptions of the VR buffet in terms of: (a) how natural was their overall experience in the VR buffet; (b) how much their final selection in the VR buffet represented a lunch that they would select and eat/drink on an average day; and (c) how much their final selection represented the lunch that they would select and eat/drink if the same food selection was available?

**Method: Participants.** Participants included 35 individuals (18 to 25-years-old) recruited from a university campus ( $M$ age = 20.49,  $SD$  = 2.17; 20 females; 15 self-identified as ethnically White, 11 as Asian, 5 as Black, 1 as Hispanic/Latinx, 1 as Bi-/multi-racial and 2 as "Arab/Arabic").

**Procedures and Measures.** Participants completed an online survey assessing demographic, psychological, and sociocultural variables. Participants were then randomly assigned to either the VR session or the RW session at the campus dining hall during lunch (11am-2pm). One week later, they were assigned to the other condition. Participants fasted for 4 hours before both sessions, and height, weight, and anthropometric measures were taken. Participants were instructed to make food selections as they would normally and there were no restrictions on the number of plates/bowls/cups they could use.

The VR buffet was modeled after the campus dining hall, and virtual models of foods were created using photogrammetry (Figure 1). Participants set their plates, bowls and cups on the virtual counter when they were finished.

To make virtual food a useful proxy for real food, we derived the number of calories from the food in both environments. Research assistants calculated the weight of the portions based on real food volume and food type. The number of calories and other nutritional information for each serving of each food type was then calculated by weight and food type.

At the end of both sessions, participants responded to three questions using a visual analogue scale (0 = Not at all to 100 = Completely); "(1) How natural was the overall experience?", "(2) How much did your final selection represent the lunch that you would select and eat/drink on an average day?", and "(3) How much did your final selection represent the lunch that you would select and eat/drink if the exact same food selection was available?"

**Results:** Participants' total serving amounts in the VR and RW buffet environments were significantly and positively correlated, such that individuals who selected a higher total number of grams in the VR buffet also selected more total grams in the RW buffet,  $r = .68$ ,  $p < .001$ . The total number of calories selected in the VR and RW food buffets were also significantly and positively correlated,  $r = .60$ ,  $p < .001$ . Moreover, significant positive correlations for high versus low density foods and macronutrients across both sessions were also found (Table 1).

Participants' ratings on "How natural was the overall experience" in the VR buffet environment ( $M = 70.97$ ,  $SD = 20.92$ ) were high and did not differ from their ratings in the RW buffet environment ( $M = 77.37$ ,  $SD = 26.13$ ),  $t(34) = 1.87$ ,  $p = .070$ , indicating that they perceived their VR experience as natural. Moreover, the final selections in the VR buffet were also rated as highly representative of what participants would select and eat on an average day ( $M = 84.11$ ,  $SD = 15.92$ ) and if the exact same food selection was available ( $M = 91.29$ ,  $SD = 11.00$ ).

**Discussion:** Results revealed that our VR food buffet is a valid tool to study food selection behaviors. The strong correlations between the food selection variables in the VR and RW phases

indicate that the nutritional content of participants' food selection in the VR environment is related to that of their RW food selection made one week apart, even when the available foods varied across contexts. These findings advanced the ecological validity of our VR buffet beyond previous food validation studies, in which food selection and related constructs were usually restricted to the same foods in VR and RW and assessed on the same day. Future integration of different measurement modalities (neurophysiological, behavioral, and survey) will allow for the identification of processes contributing to poorer food-related cognitions and behaviors among emerging adults and inform the development of programs to promote better health outcomes.

*Charissa Cheah (University of Maryland, Baltimore County)*

### Parent Teleconsultation to Improve Child's Feeding Outcomes

**Introduction:** Food selectivity and other feeding problems are common among youth with autism spectrum disorder (ASD) (Gal, Hardal-Nasser, & Engel-Yeger, 2011). With the high prevalence of ASD, there is an increased focus on identifying effective and efficient interventions to address the various behavioral needs with children with ASD, particularly with parents in the home. Blumberg and Kayser (2000) provided three recommendations to support parents with children with challenging behavior: (1) provide home-based services, (2) expand the unit of analysis from the child to the family, and (3) teach professionals to build collaborative partnerships. Parent-based interventions have been effective to target a variety of topographies of challenging behavior, including pediatric feeding disorders (Bloomfield, Fischer, Clark, Dove, 2018). Unfortunately, there are many barriers to providing these services in homes, particularly in remote and underserved communities (Clark Fischer, Lehman, & Bloomfield, 2019). Recent technology in videoconferencing has produced viable alternatives to meet these goals of providing services in the home. Videoconferencing has been shown to be effective and acceptable as a modality to conduct a variety of behavioral health services across settings and presenting problems (Bloomfield, Lehman, Clark, & Fischer, 2019; Fischer, Clark, Askings, Lehman, 2017). The purpose of this study was to evaluate a teleconsultation framework for providing effective behavioral feeding interventions in the home to a child with ASD. The following research questions were evaluated: (1) Is there a functional relationship between a parent-provided feeding intervention and an increase in the level of target behaviors?; (2) Can behavioral skills training be used to train a parent with ASD to implement the feeding intervention as measured by high levels of treatment integrity?

**Method: Participants.** Participants included a child with a feeding disorder and ASD, her mother, and the clinician. The child was a 5-year-old Caucasian female who had been reported a "picky eater." She demonstrated a restrictive variety of solid foods consumed orally. The therapist was a graduate student who was supervised by a BCBA who had training in pediatric feeding disorders.

**Setting.** The study occurred in two settings—the interdisciplinary pediatric feeding disorders clinic and virtually, through telehealth in the client's home.

**Hardware and software.** The therapist used a laptop to conduct telehealth sessions. The parent used various personal technologies with internet access, webcam, and microphone throughout the study. The telehealth sessions were conducted, recorded, and stored on HIPAA-compliant platforms. The parent sent data collected to the

therapist using email or text message, and the therapist then stored the data from the parent within the secure cloud-based storage platform.

**Experimental Design.** This study used a single-case research design methodology. Specifically, a series of changing criterion designs were conducted across approximations of food consumption target behaviors (i.e., touch, hold, smell) as well as within the consumption of foods (i.e., number of bites).

**Baseline.** During baseline, the caregiver presented the target foods and the clinician recorded how the child interacted with each of the target foods. Potential interactions were broken down into the following options: touch, pick up, bring to lips, contact tongue, bite, chew, and consume. Each food was presented and the parent instructed the child to engage in each approximation—there was no consequence for noncompliance during baseline.

**Intervention.** The intervention consisted of successive approximations and differential reinforcement to increase approximations of independent self-feeding of previously non-preferred foods. After the child complied with the instruction, she was given one minute of a preferred activity or a bite of a highly preferred food. The child was required to perform each action 5 times at 80% accuracy or higher before receiving a reinforcing object, food, or activity. The caregiver implemented sessions approximately 3 times a week independently and met with the clinician once over teleconsultation once per week. After each session, the parent sent data to the clinician, who provided feedback, as necessary.

**Response measurement and Interobserver Agreement.** The frequency of compliance with the feeding demand was recorded by the caregiver during independent sessions and the clinician during the teleconsultation sessions. During teleconsultation sessions, the caregiver and the clinician collected data simultaneously. Therefore, inter-observer agreement was completed for around 20% of total sessions.

**Results and Conclusions:** Following teleconsultation, the child demonstrated increased consumption of the target non-preferred foods. This modality was effective at training a parent to implement the intervention with high levels of fidelity.

*Bradley Bloomfield (University of Alabama)*

Lafayette Park



## Technology in Tandem: Characteristics and Effects of Joint Media Engagement in the Digital Age

### SYMPOSIUM

Digital media is pervasive in children's lives: Two- to four-year-olds spend over 2 hours a day with screen media (Common Sense Media, 2017). Media use is typically thought of as a solitary activity, but, especially for children, engaging with media can be social. "Joint media engagement" refers to children and adults using media together and can include activities as diverse as co-viewing a television show, playing a digital game together, or a parent reading an e-book to a child (Takeuchi et al., 2011). Indeed, 84% of parents report watching TV with their child at least some of the time and 63% report at least sometimes using apps on a mobile device with their child (Common

Sense Media, 2017). Joint media engagement also encompasses a range of levels of involvement, from simply using media together to having in-depth conversations about the media during use. The papers in this symposium each take a unique approach to addressing the characteristics and effects of joint media engagement. Our discussion will consider the findings as a whole and their implications for families trying to make smart media choices.

Paper 1 examines the role that a co-viewer has on low-SES preschoolers' attention to novel words and word learning from an educational video. Children viewed two videos that included novel words and definitions; one video was viewed with an adult co-viewer and one was viewed alone. Children's attention to the video was assessed using eye-tracking technology and word learning was assessed using both receptive and expressive word identification tasks. Results showed that attention to the areas of the screen where the novel words were represented was higher when children had a co-viewer compared to when they watched alone. Results from the learning task are forthcoming and will be included in the symposium presentation.

Paper 2 explores the use of an e-book-based dialogic reading training designed to promote parents use of strategies to get children actively engaged with the book's content. Researchers created an e-book with a virtual character who modeled these dialogic reading techniques. Over a two-week home-based intervention period, preschoolers and their parents were randomly assigned to read the training e-book, to read a basic version of the e-book without the training character, or to receive both e-books and choose which to read. After the intervention, families read a different e-book in the lab and their interactions were coded for parent and child talk. Results showed that parents who had previously read the training e-book produced more language and more diverse language than parents who had read the basic e-book, with parents who chose between the e-books falling in between.

Paper 3 asks how the emotional experience of reading with a parent compares to reading an e-book independently and how the emotional experience of reading an e-book compares to reading a traditional book. Preschoolers were randomly assigned to read a traditional book with their parent, an e-book with their parent, or an e-book independently. Results showed that children's physiological arousal (measured using e4 bracelets) was highest when reading a traditional book with a parent and lowest when reading an e-book independently. Facial affect coding showed that children expressed more happiness when reading an e-book with a parent and more interest when reading a traditional book with a parent, with independent e-book reading being coded as lowest for both.

These papers highlight the diverse characteristics and effects of joint media engagement during the preschool years. Paper 1 shows that simply having a co-viewer during media use can affect how children process novel words, potentially influencing their learning. This finding suggests that although encouraging parents to actively mediate children's media use may benefit learning, even the presence of an adult can promote attention to relevant educational information. Paper 2 provides important insight into the role that media itself can have in promoting parent-child interaction during media use. Importantly, the effects generalized to a new context, suggesting that the training e-book had a meaningful effect on parents' internalized strategies for joint media engagement during e-book reading. Finally, Paper 3 finds that children experience a different emotional experience when reading with their parent compared to reading an e-book independently. This suggests that reading e-books independently may

cause children to miss out on valuable emotional experiences that come from shared book reading.

This symposium brings together diverse perspectives on joint media engagement. Speakers include emerging and senior scholars from backgrounds in both psychology and education. Our discussant is the Director of Research for a non-profit focused on education and advocacy around children and media. This group is highly aligned with the multidisciplinary nature of the conference and the diverse audience of scientists, practitioners, and industry professionals.

*Chair: Rebecca Dore (The Ohio State University)*

### **Scaffolding Attention and Partial Word Learning Through Interactive Co-Viewing of Educational Media**

This study was designed to examine how co-viewing contributed to low-income children's word learning in digital stories. Using eye-tracking technology, we sought to determine the degree to which such adult mediation in a typically-paced digital story might affect children's attention and its relation to children's partial word learning. Specifically, the purpose of our study was to address the following questions: 1) How does co-viewing affect low-income children's attention to novel words? 2) Are there differences between low-frequency and high-frequency words? 3) How does co-viewing affect word learning? Is there a differential effect for low-frequency and high frequency words? Participants We recruited 83 preschoolers (Mage 4.3 years, SD = .37) from two Head Start centers located in high poverty areas in a large urban city. The sample was diverse: 29% African-American, 49% Hispanic, 18% West Indian, and 4% Asian or biracial; 55% were female. Standardized receptive language skills, measured by the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 2007) averaged 79.64 (SD = 15.76), more than one standard deviation below the norm. Research Design We used a within-subjects design to examine the effects of co-viewing educational media stories on children's word learning. In this within-subjects design, each child viewed two stories, one with an adult co-viewer, and the other without, in a counterbalanced approach. Word learning in each condition was compared for each individual participant. In both conditions, children viewed the video on a computer equipped with eye-tracking technology to examine their attention throughout the programs. We selected two full-length (9.5 minutes) narrative stories from the educational media program, *Peep and the Big Wild World* (produced by WGBH). One story episode focused on plants and the other, on shapes. To measure how word frequency might affect word learning, we replaced the audio track of both programs with an adapted script that incorporated 8 vocabulary words per episode. Following the plot line of the original scripts, graduate students from the educational theater program took on the roles of the characters and the narrator in the program, providing an initial ostensive definition and additional repetitions of the target words. Half of the words in each video were repeated at a lower rate (3-4 times), while the other half, at a higher rate (8-11 times). Measures Before viewing. Screening Pretest Measure. We administered a screening pretest measure to assess children's prior knowledge of target words. Cronbach's alpha= 0.52. Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 2007). We administered the PPVT to examine children's overall receptive language skills. Reliability was 0.91. While viewing. Children viewed the programs from a computer connected to an eye-tracking device. Our system uses a binocular tracking method, which allows for increased head movements. Tobii

Studio Professional 3.0 software was used for stimuli presentation and data processing. After calibration, children would then view the program, with the research assistant able to follow the child's eye movements and behaviors using the live view on the second monitor. Post-viewing assessments. Receptive Word Identification: Similar in format to the PPVT, children were shown four images and asked to point to the target word. There were 16 items per assessment, for a total of 32 items across the two videos. Cronbach's alpha= 0.61. Expressive Word Identification. Similar in format to the EOWPVT, children were shown a screenshot of each target word, and asked: "What is this?" Correct responses of the exact word (e.g. no synonyms accepted) were calculated for each video. There were 8 items per assessment for a total of 16 items across both videos. Cronbach's alpha= 0.69. Results Results of our analysis indicate that co-viewing influenced children's attention to the target words. The results of our 2 x 2 repeated measures ANCOVA, with co-viewing condition and word frequency as within-subjects independent variables and age in months as covariate, reported a significant main effect of condition,  $F(1, 69) = 49.33, p < .001, \text{partial } \eta^2 = .417$ . There was also a significant main effect of word frequency,  $F(1, 69) = 11.64, p = .001, \text{partial } \eta^2 = .144$ . However, there was no significant effect of the covariate or any significant interactions. These results indicate that the attention to target words was greater in co-viewing condition than in the non-interactive condition. At the same time, it also suggests that word frequency makes a difference: Regardless of condition, children were more likely to attend to words if they were named more frequently. Results from the learning task are forthcoming and will be included in the symposium presentation.

*Kevin Wong (New York University)*

### **Home Use of a "Dialogic Reading" e-Book Promotes Parent-Child Talk with Another Book**

Dialogic reading, an intervention designed to promote parent-child talk during reading, leads to preschoolers acquiring more vocabulary during reading (Whitehurst et al., 1988; Mol, de Jong, & Bus, 2008; Sloat, Letourneau, Joschko, Schryer, & Colpitts, 2015). However, despite a 30-year research history, dialogic reading techniques have not been widely adopted by parents reading with preschoolers. One barrier may be a lack of accessible mainstream tools for parents to learn these techniques. In a prior study, we showed that a virtual e-book character who modeled dialogic reading techniques could increase parents' use of those strategies during reading (Troseth et al., 2019). However, the study was limited in ecological validity, as parents read the training book in the lab, and talk was measured during training. In the current study, we investigated whether parents could be successfully trained by using the e-book with their child at home, and whether parents would learn skills transferrable to new books that did not include the virtual training character. To date, 42 parents and their three- and four-year-old children have participated in one of three conditions (final anticipated N=75). In all conditions parents and children visited the lab and read a print version of *Peg + Cat's The Big Dog Problem* to provide a baseline measure of parent-child talk during shared reading. Parents then read *The Big Dog Problem* e-book with their 3- or 4-year-old child at home ten times over two weeks. In the control condition, parents and children read this same e-book without the training character. In the training condition, they read a version of the e-book in which the training character modeled simple questions in week 1 and a version of the same e-book with

more complex questions in week 2. In the choice condition, all three e-book versions were available for the families to choose between each time they read. At the end of the two-week reading period, parents and children returned to the lab and read a different e-book, *The Election Problem*. Use of the training e-book successfully promoted parent talk while reading the new book, as indicated by ANCOVAs controlling for pre-test talk, and post hoc comparisons with Sidak corrections. Parents in the training condition provided significantly more utterances ( $M = 72.47$ ,  $SD = 36.90$ ), words ( $M = 405.07$ ,  $SD = 226.69$ ), and unique words ( $M = 119.47$ ,  $SD = 42.81$ ) than parents in the control condition did (utterances  $M = 33.38$ ,  $SD = 21.74$ ,  $p = .002$ ; words  $M = 159.92$ ,  $SD = 109.11$ ;  $p = .001$ ; unique words  $M = 70.38$ ,  $SD = 36.21$ ,  $p = .003$ ). The training e-book also promoted child talk; children in the training condition used significantly more utterances ( $M = 26.33$ ,  $SD = 16.69$ ) than children in the control condition ( $M = 10.77$ ,  $SD = 7.46$ ,  $p = .04$ ). Parents and children in the choice condition had means more similar to those in the training than the control condition, perhaps partially due to large variability (parent utterances  $M = 70.64$ ,  $SD = 51.05$ ; words  $M = 339.71$ ,  $SD = 252.23$ ; unique words  $M = 108.57$ ,  $SD = 57.95$ ; child utterances  $M = 27.07$ ,  $SD = 21.98$ ). Coding of conversational turns and the content of parent and child language, including the use of specific dialogic reading techniques, is ongoing and will be ready to present at the conference. Our preliminary results indicate that a training e-book used by families at home can promote increases in parent and child talk when reading new e-books together. Exposure to more language (Weisleder & Fernald, 2013), more diverse language (Hoff, 2003), and more back-and-forth conversation (Romeo et al., 2018) at home has been linked with children's language growth and skill. This is the first step in showing that parents can learn transferrable skills with the potential to impact children's language development from an e-book with a virtual character who models dialogic reading techniques. The training e-book has the potential to bring dialogic reading techniques into mainstream practice by providing an easily deployable training tool that parents and children can use together in their own homes during regular shared reading experiences.

*Gabrielle Strouse (University of South Dakota)*

### **Emotional Experiences During Parent-Child e-Book Reading: Physiological Arousal and Facial Affect**

Shared storybook reading is very common: ninety-one percent of parents with children under 6 report reading books aloud (Scholastic Inc. & YouGov, 2015). However, with the increasing presence of mobile devices in homes, over 70% of parents report sometimes or often giving their child an e-book to read alone when they are busy doing something else (Vaala & Takeuchi, 2012). Children do comprehend some content from e-book audio narration, although comprehension is better after parent reading (Dore et al., 2018). Parent-child reading may also have emotional advantages over independent e-book reading; bonding with children is one of parents' most important goals for parent-child reading (Audet et al., 2008). Furthermore, there may be emotional differences between reading traditional books and e-books. Yuill and Martin (2016) found that observer-rated interaction warmth was lower when 7- to 9-year-olds and their mothers read an e-book compared to a traditional book. In the current study, we explore these different reading experiences through facial affecting coding, self-reported emotion, and physiological

arousal. Our preliminary data consists of 37 four-year-olds (Mage = 52.80 months; 23 girls) and their mothers. Children were randomly assigned to read the same storybook in one of three conditions: reading a traditional book with parent ( $N = 12$ ), reading an e-book with parent ( $N = 13$ ), or listening to an e-book independently ( $N = 12$ ). Children wore an Empatica E4 multi-sensor to capture electrodermal activity (EDA), an index of sympathetic arousal. For each minute of baseline activity—a series of mindfulness games (Boguszewski, & Lillard, 2015)—and reading, we examine the number of skin conductance responses (SCRs) as a measure of focused attention or interest. To assign affective valence to child EDA, videos were coded with a microanalytic coding system using affect expressions to capture the percentage of time children displayed different emotions during reading (adapted by Roben, 2018 from Cole et al., 2007). Facial, affect, and postural cues are based on research on basic emotions (e.g., Ekman & Friesen, 1975; Izard, 1979). We also used self-report measures at pre- and post-test to assess how happy, excited, sad, sleepy, and nervous the child felt and how much they liked reading the book. Results showed that children had significantly more SCRs when reading a traditional book with their parents ( $M = 2.99$ ,  $SD = 0.97$ ) than when listening to an e-book independently ( $M = 1.51$ ,  $SD = 1.01$ ) after controlling for baseline,  $p = .016$ ,  $d = 1.49$ . SCRs for children who read an e-book with their parent fell in between ( $M = 2.25$ ,  $SD = 1.02$ ) parent traditional ( $p = .644$ ,  $d = .75$ ) and independent e-book ( $p = .620$ ,  $d = .73$ ). Children's facial affect was coded as happy during more of the session when they were reading an e-book with a parent ( $M = 14.3$ ,  $SD = 10.7$ ) compared to when they were reading an e-book independently ( $M = 2.5$ ,  $SD = 6.5$ ),  $p = .014$ ,  $d = 1.33$ . Parent traditional book ( $M = 7.0$ ,  $SD = 8.2$ ,  $n = 10$ ) fell in between parent e-book ( $p = .175$ ,  $d = 0.77$ ) and independent e-book,  $p = .682$ ,  $d = 0.61$ . Children's facial affect was coded as interested during more of the session when they were reading a traditional book with their parent ( $M = 9.3$ ,  $SD = 5.0$ ) than reading an e-book independently ( $M = 2.0$ ,  $SD = 2.1$ ),  $p = .008$ ,  $d = 1.90$ . Parent e-book ( $M = 6.4$ ,  $SD = 7.3$ ) fell in between independent e-book ( $p = .213$ ,  $d = 0.82$ ) and parent traditional book ( $p = .598$ ,  $d = 0.46$ ). On the self-report measures, children reported being significantly less nervous after reading a traditional book with their parents ( $M = -1.36$ ,  $SD = 1.57$ ) than reading an e-book independently ( $M = 0.33$ ,  $SD = 1.56$ ),  $p = .023$ ,  $d = 1.08$ . There were no significant differences for the other self-reported emotions or for liking the book reading activity. These findings showed that reading with a parent is a different emotional experience than reading an e-book independently. Reading with a parent was associated with greater physiological arousal and more positive emotion than reading independently. Children may have been happier and more excited during the e-book reading because e-books are perceived as more novel and fun. On the other hand, children may have been more interested when reading a traditional book with their parent because traditional books are more familiar and more associated with learning leading to deeper engagement and focus. Overall, these findings suggest that reading e-books independently may cause children to miss out on valuable emotional experiences that come from shared book reading. Furthermore, reducing the positive valence of typical reading experiences may have the long-term effect of making reading less attractive relative to other activities that do evoke positive emotion.

*Daniela Avelar (University of Delaware)*

*Discussant: Michael Robb (Common Sense)*

12:15–1:30 P.M.

## Lunch on Your Own

---

## KEYNOTE ADDRESS

1:30–2:45 P.M.

Independence Ballroom

# The Digital Revolution: The Potential Promise and Ethical Perils in Research

Digital tools including apps, wearable sensors and social network platforms offer unprecedented opportunities for research. However, this rapidly evolving landscape is outpacing existing regulatory structures for protecting research participants. In addition, new actors in the scientific community, including technology companies and citizen scientists are not bound by the rules used by traditional academic researchers to guide responsible and ethical research.

Dr. Nebeker will describe how technologies are being leveraged to capture personal health data for research drawing attention to nuanced technical and ethical aspects that require careful consideration during the study design phase. She will underscore the important role of funding agencies, policy makers, editors, researchers and ethicists in creating the infrastructure necessary to allow digital health research to flourish.



## CAMILLE NEBEKER, EdD, MS

### Director

*Research Center for Optimal Digital Ethics  
(ReCODE.Health)*

### Associate Professor of Behavioral Medicine

*Department of Family Medicine and Public Health,  
School of Medicine, UC San Diego*



3:00–4:15 P.M.

## Concurrent Sessions 5

Independence BC

### Children and Technology

#### PAPER SESSION

##### Developing a Mathematics Application to Improve Children's Magnitude Comparison Skill

Problem Statement. Information and Communication Technologies (ICT) have been reported as an emerging influence in every aspect of our lives, including living, learning, and interacting with others (Mera et al., 2019). At present, children are using technological devices which expand their knowledge in mathematics (Plowman et al., 2012). There is an increasing use of digital tools in home settings which further boost the design of new digital applications (apps) to improve children's learning. Regarding math deficit issues, few apps have been developed considering domain-specific deficits such as processing speed, magnitude representation (MR). Particularly, poor MR is among the most common that are thought to underlie math difficulties (Chan, 2014). Children with these deficits demonstrate impaired performance on magnitude comparison tasks, such as comparing Arabic numerical digits, 2 VS 4 (Bugden & Ansari, 2011; Mazzocco et al., 2011). The present paper explores the design and development of a mobile math app with the purpose of implementing it as a math learning tool for young children, ages 5-8. The objective is to investigate the effectiveness of a mobile math app through several hours of play in a home context. In particular, the math app has been designed for improving children's MR ability. Procedure. Our research team and a game developer company (<https://www.zapzapmath.com>) have worked together to design and develop the app for smartphone devices and to adapt it to different operating systems (i.e., Android, IOS). The MR app consists of three segments of numerical comparison activities, i.e., a symbolic comparison between two Arabic digits, a non-symbolic comparison between two set of numerical dots and a mixed comparison between Arabic digits and dots. The app has 27 levels of difficulty to accomplish, consisting of 798 pairs of numerals (i.e., each segment has 266 pairs). Each level of difficulty is determined by the ratio between the number pairs (e.g., 2:3) During the game, the numerical ratio of the pairs is manipulated to present the smaller, easier ratios first, followed by increasingly larger ratios to keep the player motivated (Hawes et al., 2019). The ratio (small to large) of the pairs follow a range from 2:3 to 6:7 ratios as developing children can discriminate that ratio range (Halberda & Feigenson, 2008). The numbers range from 1-100 following the Canadian Education System; Grade 1 students are taught to count until 100 (e.g., Lester B. Pearson Board math curriculum). The targeted players (n=25) of the app, will be children from Grade 1 through 3 with math difficulties, which will be identified based on their performance on standardized math measures (i.e., KeyMath-3 Diagnostic Assessment Tool, Connolly 2007). Children will play the math app for 6.25 hours (15 minutes per day, 5 days a week for 5 weeks) in their home. Before starting the game, four trials will be included to confirm that the children understand the task. Children's correct response and reaction time will be recorded on the log file in a detailed live data (such as duration of play, correct answer etc.).

Following that, children will complete Key-math 3 form B to identify changes in their math performance. In addition, they will complete an MR task in both the pre and post-test sessions. Analysis. Repeated measure Analysis of Variance (ANOVA) will be used to measure and compare between the duration and quality of play (e.g., errors during play) on 5 weeks. To investigate whether MR app improves children's math performance over time, a multivariate analysis of variance (MANOVA) will be used to compare their score on both Keymath-3 and MR task (pre-test and post-test) by type of MR app segment (i.e., symbolic, non-symbolic and mixed). Results. The study will explore whether the MR activities help to improve children's domain specific cognitive ability. Moreover, it will explore the function and inter-relation of different types of MR numerals (i.e., symbolic, non-symbolic, and mixed) involve in math cognitive mechanisms which is a vague concept and area of debate till today. Practical Implications & Conclusion. If effective, the math app will be an easy to implement digital home numeracy practice (DHNP) that parents can use to help their children's fundamental numeracy ability.

*Sabrina Shajeen Alam (McGill University)*

##### Digital Technologies and Mediation: A Vygotskian Perspective

In the first half of the 20th century, a founder of cultural-historical psychology Lev Vygotsky proposed a framework to understand cultural development of human higher psychological functions (Vygotsky, 1978). As one of the core concepts, he introduced a concept of cultural mediation as acquisition of new cognitive "tools" from adults within a so-called "zone of proximal development". According to Vygotsky, higher psychological functions are first divided or shared between a child and an adult in such a manner, that what the child cannot do on his/her own, he or she can perform together with an adult who meanwhile introduces cultural tools which mediate cognition and performance (e.g. numbers for counting, mnemonic tools for better remembering, etc.). Then these "mental tools" are being internalized to become tools of an individual mind.

However, the world around us and in particular technologies change faster than one could imagine, providing new kinds of support for our cognitive functions (memory, spatial navigation, attentional orienting, and many others) and emphasizing an urgent need to investigate a developing mind in the developing culture (e.g. Cole, Packer, 2016).

I would argue that the modern digital technologies cause two remarkable reversals in the course of cognitive development as outlined in Vygotsky's cultural-historical psychology.

First, for centuries, the general trend in both individual cognitive development and the cognitive evolution as a whole was from external to internal tools of the mind (e.g. from real knots to mental notes, or from marks on the wood to mental calculations). But now



our higher psychological functions become mostly externalized again due to the use of new digital tools, such as reminders, web search instead of memory search, marked keywords which guide attention, etc. In other words, the humanity moves back from internalization to “new externalization”, with digital tools becoming an integral part of our cognitive functioning. In philosophy, this phenomenon has been described as Extended Cognition (Clark, Chalmers, 1998). From the psychological viewpoint, this means reconstruction of the system of higher mental functions through the digitally mediated activity (cf. Kaptelinin, Nardi, 2009).

Second, the zone of proximal development is also being transformed by digital devices, because kids now mostly master them on their own, out of the joint activities with adults, and the adults don't even need to share their skills and experience. Of course, the devices do incorporate certain cultural practices, as well as any other cultural artifact, such as a fork or a spoon which force a baby to use them in a certain way, transforming his or her movements. However, with the new digital technologies, the adult is no longer necessary as an instructor, and the child's attention, memory, cognition, and activity are being structured and organized by the interaction with the device itself. The potential problem is that the developmental trajectory of “digital natives” becomes less predictable for the previous generation, diverging from the “inter-individual to intra-individual” general trend, described by Vygotsky. The new gadgets become “material signs” for the developing generation, just as primitive material tools, such as choppers, at the earlier stages of human evolution, which embodied specific affordances and prompted the further actions on them (Malafouris, 2013). This means that the very concept of the zone of proximal development requires reconsideration, so that it could incorporate not only human-human (child-adult), but also human-computer interaction.

To sum up, on the one hand, new digital technologies challenge the cultural-historical approach towards cognitive development and towards the understanding of the evolution of human cognition, but, on the other hand, the constructivist nature of this approach and the concept of cultural mediation might provide new insights into the nature of extended cognition.

*Maria Falikman (Higher School of Economics)*

---

Farragut Square

## Work and Management in a High-Tech World

### PAPER SESSION

#### **“Connecting Through Feedback”: Creating an Organisational Culture for Collaboration, Shared Problem Solving and Innovation in ICT**

The problem: Openness to feedback is an attitude and skill fundamental for the current and next generation of computer practitioners. More and more they are expected to work in interdisciplinary teams, engage in collaborative design processes and take into consideration the moral implications of their work. But where does that openness to feedback come from and what are our academics modelling for our students? This poster presentation describes a long-term organisational culture change intervention in a university information technology department (school). The intervention, “Connecting Through Feedback”, was designed to shift a pre-existing norm of academics

avoiding discussing student feedback about their learning and teaching practices, thus creating greater opportunities for collaboration, shared problem solving and innovation.

In early 2014, it became evident to a school of Information and Communication Technology (ICT) leaders that the university-wide initiative to implement student success and retention teams were not forming. Team members reported an unwillingness of some colleagues to share feedback about their students and engage in constructive conversations about their teaching practices. In response, the school introduced a school based reflective practice process, known as Connecting Through Feedback, led by a psychologist/internal consultant (from a central academic development unit) and mid-career academic from the school. This process was designed specifically to shift norms related to how school-based academic staff interact with students, their peers within the school, and faculty and university learning and teaching experts. The process has been running since 2014 and consists of a series of workshops (two a semester) that create an opportunity for colleagues to come together to reflect on their course feedback and student data, support each other to analyse the information and develop a plan for what they would do. It is managed within the school and facilitated by academics.

Intervention: The Connecting Through Feedback process extends the scholarly practice of individual critical reflection (Potter & Kustra, 2011) using a group setting to create the conditions for double loop learning (Argyris and Schön, 1974). Wilson (2011) describes double loop learning in scholarly practice as critical reflection which stretches and challenges assumptions and conceptions, in order to improve as well as innovate. As these team based reflective practice cycles continued across semesters, participants reflected on what worked and what needed further consideration. The scholarly teaching cycle repeats as new hot data becomes available. Students participate in each cycle through giving feedback to their lecturers and then receiving information about what is planned in response, (see Figure 1.). Overtime the workshops have evolved to suit the needs of the group, for example the participants may choose to focus on a specific aspect of practice, such curriculum design.

The Connecting Through Feedback process is an iterative series of two action learning cycles (Argyris & Schön, 1974) per semester. The complete iterative Connecting Through Feedback activity is outlined in Figure 2 (Lizzio & Jeremijenko, 2014).

In the first Action learning cycle, shown in Figure 1, the four action learning phases are as follows:

Phase 1: Experience: Feedback on week 1-4 teaching is sought from students during week 4 using the Closing the loop model detailed previously.

Phase 2: Collect/reflect: In week 5, teaching staff collate and analyse the data gathered from Phase 1 using a worksheet detailing the analysis and key issues.

Phase 3: Analyse/theorise: In week 6, a 1 hour workshop is held in an environment that is safe and promotes double loop learning (Argyris & Schön, 1974). Participants create an action plan based on their discussions and a whole group debrief is held. An evaluation is also performed.

Phase 4: Plan: In week 7, staff complete the action plan begun in week 6 and disseminate the list of issues and their corresponding action plan to students to close the feedback loop.

Results: The intervention used an action research approach, enabling evaluation data to be collected from multiple sources at multiple times (Kember, 2000). Methods include: survey, organisational

performance indicators, semi-structured interviews and research team reflections. Preliminary evaluation results were positive and resulted in the leaders of intervention receiving a competitive citation for learning and teaching practice from the university. Final evaluation results will be published after a follow-up survey is administered in June 2019.

Implications: The implementation of the intervention was underpinned by a transformational change approach (Venema & Jeremijenko, 2017) to develop a culture where gathering feedback and freely discussing it with other academics is seen as both normal and valuable. Results indicate it has created organisational and team learning environments (Senge, 1990), where double loop learning is part of the practice rather than a one-off intervention. In this transformed environment, colleagues share their practice and teaching innovations with others and best practice could evolve through shared experience.

*Samantha Carruthers (Griffith University)*

### **Virtual (Freedom From) Reality: When and Why Managers Prefer to Monitor Subordinates Through Avatars**

The present research examines the conditions under which managers prefer interacting with their subordinates virtually, via computer avatars, as opposed to through traditional face-to-face conversations. Two experiments tested the prediction that managers prefer to use technological interaction tools to interact with their subordinates when they are concerned about negative social evaluation. This prediction is based on the idea that technology acts as a psychological buffer, allowing people to avoid face-to-face interactions in uncomfortable social contexts. Using computer avatars (graphical computer representations of humans), we conducted two experiments to test this prediction. Results supported our prediction, indicating that participants in managerial/leadership roles avoided face-to-face interactions and preferred virtual interactions in situations where they were concerned about negative social evaluation. Specifically, in Study 1 (N = 101), participants in a managerial role, who were asked to monitor their subordinates frequently, preferred interacting with their subordinates via computer avatars. In Study 2 (N = 196), participants, in their roles as team leaders, preferred interacting with their team members via computer avatars and this effect was mediated by reduced concerns about negative evaluation. We also explored how personality variables influenced this effect. We discuss the theoretical implications of these findings and highlight future directions for research on the psychology of technology.

*Roshni Raveendhran (University of Virginia)*

### **I Didn't Agree to These Terms: Electronic Performance Monitoring Violates the Psychological Contract**

Introduction and Statement of Problem: Electronic performance monitoring (EPM) is the electronic set of methods for collecting, recording, storing, or analyzing employee behavior (e.g., video surveillance, GPS location tracking, internet monitoring; Alge, 2001; Stanton, 2000). Reactions to this technology differ widely. Monitoring can elicit perceptions of privacy invasion (Moorman & Wells, 2003; Stanton, 2000) and unfairness/injustice (McNall & Roch, 2009; McNall & Stanton, 2010), yet little is known about the relationship

between expectations of EPM and reactions. We build upon past work (Tomczak, Willford, White, & Behrend, 2018) to approach this problem from both job attitudes and job characteristics perspectives.

The psychological contract is the unstated expectations of the employee-employer relationship, such as satisfactory performance and mutual respect (Rousseau, 1995). Psychological contract violation (PCV) is an emotional reaction to the organization's failure to uphold this contract, which may lead to counterproductive work behavior (Bordia, Restubog & Tang, 2008) and turnover (Jensen, Opland, & Ryan, 2010; Morrison & Robinson, 1997; Robinson & Rousseau, 1994). Employees may perceive EPM as a signal of distrust, which may result in PCV (H1).

Job characteristics may also provide insight into monitoring expectations (Hackman & Oldham, 1976). Autonomy is an employee's perceived ability to have control over their work behaviors (Morgeson & Humphrey, 2006). Higher levels of perceived control relate to greater involvement and less withdrawal (Spector, 1986). Because EPM is often used to control behavior, the EPM-PCV relationship may be stronger when perceived autonomy is low (H2).

Job complexity is the extent to which job tasks are complex, difficult to perform and require a variety of skills to complete (Morgeson & Humphrey, 2006). Complex work is more challenging and mentally demanding, and thus may require more effort (Joo & Lim, 2009). According to the job demands-resources model, however, employee autonomy is necessary for motivating employees to apply full effort in complex jobs (Bakker & Demerouti, 2006). EPM may negatively affect employees' desire to exert maximum effort in complex jobs if it restricts an employee's autonomy (H3).

Methods and Procedures: Participants were recruited from Amazon Mechanical Turk (MTurk). Only individuals with high-quality data (passed attention checks, complete data) were included, resulting in 154 participants. The sample was 59% male and 81% white, with 61% of participants within the ages of 18 to 35.

EPM amount was measured using a checklist to indicate all methods in which participants were monitored at work. Participants completed measures of perceived job autonomy, withholding effort, and PCV, and provided their job title. We operationalized job complexity as the sum of KSAOs listed for the job on O\*NET and the automatability of the job according to Frey & Osborne (2013), which calculated the level of creativity, social intelligence, and perception/manipulation required for each job using O\*NET.

Analyses and Results: Descriptive statistics are reported in Table 1, showing that EPM was not correlated with withholding effort ( $r=.02$ ). We used Preacher, Rucker & Hayes (2007)'s bootstrapping method for testing moderation and mediation to test H1-H3.

The indirect effect of PCV on the relationship between EPM and withholding effort was .055 (Figure 2), and the confidence interval did not include 0 (BootLLCI = .0148, BootULCI = .1009); thus H1 was supported.

The EPM-autonomy interaction was significant ( $\beta=.20$ ,  $p=.03$ ) (see Table 2 and Figure 3), supporting H2.

The interactions for job complexity ( $\beta=-.01$ ,  $p=.78$ ) (see Table 3) and automatability ( $\beta=.05$ ,  $p=.28$ ) (see Table 4) were not significant; thus H3 was not supported.

Implications and Conclusion: The exploratory study in Tomczak et al. (2018) found that reactions to EPM can be explained from job attitudes and job characteristics perspectives. Results from the current study confirm the role of job attitudes, specifically perceived autonomy and PCV, in withholding effort in response to

EPM, suggesting that EPM violates the unstated psychological contract between employee and employer. More research is needed to understand how these expectations arise and how organizations may recover from PCV once it has occurred. The moderating role of job complexity found in Tomczak et al. (2018), did not replicate in the present study, and post-hoc analyses suggest that this may be because privacy attitudes and expectations have changed over time. More research is needed to explore the importance of job complexity and incorporate other aspects of job characteristics, such as task variety or task identity. Although EPM violates the psychological contract regardless of job characteristics, these findings suggest that violation perceptions are least severe when perceived autonomy is high and monitoring is low.

*David Tomczak (The George Washington University)*

## Independence E

### Machine Learning and Education

#### PAPER SESSION

##### Driving Reading Engagement via Machine Learning

Data offers inspiring possibilities for solving large social problems around the world. This is especially relevant to the global education community where data can be leveraged to improve learning outcomes (Hilbert, 2016; UNESCO, 2017). This presentation will explore how machine learning and statistics can provide insights into reading behaviours, and hence inform programme design.

In line with its mission to proactively encourage reading, Worldreader offers free books to mobile phone owners via free mobile reading apps. During the past 9 years these apps have had approximately 11 million users from over 50 countries around the globe. And presently, approximately 1 millions app interaction records are recorded each day. The interactions include reading, searching, and browsing interactions that can provide invaluable insights.

First, we will showcase a model used to determine the reading speed ranges of the reading application users. It allows the organisation to distinguish between reading and browsing behaviour via reading speeds. Reading speed is the ratio of text interacted with, in characters, to interaction time. It is objective, consistent and trackable and can be used as a versatile measure for a variety of actionable steps, e.g., determining the readability and difficulty of texts, exploring literacy rates, etc. (See Figures in supporting documentation)

Second, we will discuss the result of profiling model used to preliminarily identify the reading profiles of our Kenya users. The model is an ensemble model of singular value decomposition, gaussian mixture model, and relies on the reading data of individuals in Kenya. This two-year research project that sought to understand the barriers and drivers for women and girls in Kenya to read using their mobile phones. The project built a collection of books for women, and resulted in a campaign for how to better encourage women and girls to read on their mobile phones.

Using these case studies the presentation will outline how big data can be leveraged to track book performance, and automate processes to build predictive machine learning models. The presentation will further demonstrate how, with further investment in these tools, we will be able to establish links between mobile reading of

e-books and improved digital literacy skills by tracking user paths and behaviour evolution through our mobile reading app. Implications for improving literacy for both school age and adult populations will be discussed.

*Catherine Ojo (Worldreader)*

##### Modeling Implicit Computational Thinking in Zoombinis Mudball Wall Puzzle Gameplay

Introduction: There is increasing evidence that games can play a significant role in promoting STEM learning for children and young adults (Authors, 2017). Well-designed games provide the opportunity to support the development and measurement of implicit learning, which is argued to be foundational for explicit learning in the classroom (Polyani, 1966). Computational thinking (CT) is important to the development of 21st century skills (Voogt et al., 2015), and may also support the inclusion of learners with cognitive differences. Previous studies have shown a potential link between cognitive differences and computational thinking, with students with autism excelling at pattern recognition (Dawson et al., 2007) and learners with ADHD demonstrating superior performance on idea generativity than adolescents without ADHD (Abraham et al., 2006). As such, assessing implicit CT through gameplay is not only valuable in its own right, but can also help educators broaden and deepen STEM experiences for diverse learners.

Learning analytics methods offer unique opportunities for providing scalable, replicable measures of implicit learning in games (Authors, 2017). In this study, we used learning analytics techniques to build automated detectors able to recognize students' implicit Computational Thinking (CT) skills and strategies from gameplay log data. Success in accurately detecting these skills and strategies represents a new way of providing formative, real-time game-based assessments of CT.

Zoombinis (Authors, 2015) is an award-winning, popular educational game that elicits implicit computational thinking. We study CT within Mudball Wall, where a large gridded wall blocks the Zoombinis' progress (Figure 1). Players must use computational thinking skills and strategies to discover the correct patterns of shapes & colors of mudballs to hit dotted cells on the wall and catapult their Zoombinis onward.

Methods: Our process of modeling implicit computational thinking includes five steps: 1) Reliable hand-labeling of CT behaviors within Zoombinis gameplay, 2.) Synchronizing labels to gameplay process data, 3) Distilling gameplay process data into features useful for measuring behaviors consistent with CT, 4) Building detectors of players' CT strategies based on human labeling, 5) Validating the detectors as formative assessments of implicit CT by comparing to performance on external pre/post assessments.

In step 1, researchers watched Zoombinis gameplay independently and refined label definitions of implicit computational thinking, until they agreed that the labeling system was an exhaustive representation of all the CT-related gameplay behaviors seen. Next, two researchers independently labeled all rounds of Level 1 play from 74 students (37 upper elementary and 37 middle school), achieving acceptable interrater reliability (Cohen's kappa = 0.733 to 0.940 across CT skills and strategies). Figure 2 shows sample labeled data.

We then built separate detectors for each human-applied label of implicit CT, trying four common classification algorithms used to predict categorical variables from a set of features: W-J48, W-JRIP, step

regression, and Naive Bayes. The goal of building these detectors is to replace the hand labeling of CT labels with an automated model that can be applied directly to data. Detectors were built using 4-fold student-level batch cross-validation, a process in which models are repeatedly built on 75% of the students and tested on the remaining 25%, to estimate model generalizability to new data. These detectors were evaluated using two goodness metrics: Cohen's Kappa and AUC ROC.

**Results and Discussion:** Table 1 presents the best performing algorithm for each CT label in Mudball Wall. While the 2D Pattern Completer and Alternating Color and Shape strategies obtained relatively low kappa values, all detectors achieved moderately high to excellent AUC values, in the range of 0.702 to 0.976. Specifically, the Implementing a Full Solution and Abstraction detectors both achieved an AUC value of 0.976, meaning that the detectors can correctly distinguish between the absence and presence of these CT skills 97.6% of the time. This is a level of quality higher than seen in many medical applications, indicating that these detectors can be used for formative assessment of students' implicit computational thinking and for informing teachers which students need learning support.

Detecting implicit computational thinking has potential implications for providing an innovative form of assessment for diverse learners who often face challenges in demonstrating their knowledge in the context of traditional assessments. Diverse learners may be more able to demonstrate what they know when learning occurs in a well-designed game, where they are able to bypass potential barriers of traditional assessments and test solutions across puzzles that do not rely on word problems. This work, therefore, represents a step towards providing implicit STEM learning assessments using learning analytics techniques toward a more effective and inclusive STEM education. The final step of our process, not yet complete, will be to apply these detectors to a broader sample and validate them against external pre-post assessments of computational thinking collected as part of a national Zoombinis implementation study.

*Ma. Victoria Almeda (TERC)*

## Independence HI



## Engaging Patients and Transforming Care

### PAPER SESSION

#### Attention Behaviours Can Increase User Engagement with Healthcare Robots

**Background:** The use of robots in healthcare environments is fast becoming a reality. However, in order for robots to be effectively employed in health applications, they need to not only be useful, but also display appropriate social behaviours. The ability to attract, maintain, and display attention is a fundamental aspect of social interactions. This research examined the effect of robot attentional behaviours on user perceptions and user behaviours in a simulated healthcare interaction.

**Methods:** This study utilised a between-subjects experimental design. Participants (N = 181) were recruited from both the University of Auckland, and wider Auckland community. After completing baseline measures, participants were randomised to one of four conditions in which a robot used forward lean, voice-pitch changes,

self-disclosure, or neutral behaviours, during a scripted interaction. Participants completed post-interaction measures and interactions were video-recorded.

**Findings:** Participants in the forward lean and self-disclosure conditions found the robot to be more stimulating than participants in the voice-pitch or neutral conditions ( $p = .03$ ). Participants in the forward lean, self-disclosure, and neutral conditions found the robot to be more interesting compared to those in the voice-pitch condition ( $p = .002$ ). Participants in the forward lean and self-disclosure conditions spent more time looking at the robot than participants in the neutral condition ( $p < .001$ ). More participants in the self-disclosure condition laughed during the interaction ( $p = .01$ ). More participants in the forward-lean condition leant towards the robot during the interaction ( $p < .001$ ).

**Discussion:** The use of self-disclosure and forward lean by a healthcare robot can increase human engagement and attentional behaviours. Voice-pitch changes did not increase attention or engagement. Research looking at voice pitch changes may need to use more pronounced, frequent changes, in order to see potential effects. It is important to program attentional behavioural into healthcare robots to increase user engagement.

*Elizabeth Broadbent (The University of Auckland, New Zealand)*

#### Empirically Examining the Reach and Impact of mHealth Apps for Suicide and Non-Suicidal Self-Injury Prevention: A Systematic Review

**Introduction:** Suicide and self-injury are important public health concerns. Individuals experiencing suicide and non-suicidal self-injury often experience barriers to treatment, including high cost, stigma, and preferences for self-help (Czyz et al., 2013). Mobile health (i.e., mHealth) interventions have the potential to expand access to mental health services (Kadzin, 2017) and have been developed for suicide and self-injury (Larsen, Nicholas, & Christensen, 2016). mHealth apps are a particularly attractive medium for treatment delivery due to their scalability and ease of uptake (Fairburn & Patel, 2017; Kazdin, 2017). Indeed, approximately 65% of smartphone users report using an app for mental health symptoms in the past month (Rideout & Fox, 2018). Despite this, limited research has examined the reach and impact of existing digital interventions. Although mHealth interventions have the potential to improve access to services, is currently unclear if and how these interventions acquire and retain users. If we understood the distribution of downloads and active users among these apps, we would better understand the dissemination of existing interventions and directions for future research. In this pilot study, we examine the downloads, daily active users, and monthly active users of publicly available apps for suicide and non-suicidal self-injury prevention.

**Methods:** We conducted a systematic review of the Apple App Store and Google Play Store and included apps designed to prevent suicide or nonsuicidal self-injury. We collected downloads, daily active users, and monthly active users from Mobile Action (<https://www.mobileaction.co/>), a mobile app analytics tool which provides estimates for app usage and retention data. We collected these metrics for the six-week period from October 1 to November 15, the same dates as our search of the commercial app stores. Independent raters also assessed each app for engagement, functionality, and aesthetics using the Mobile App Rating Scale (MARS).

**Results:** We included 91 apps in our review; 60 were available on the Apple App Store and 75 were available on the Google Play Store. Apps on the Apple App Store varied widely in their number of

downloads (Mean = 400.21, Median = 0, SD = 1094.69), daily active users (Mean = 1435.96, Median = 0, SD = 3939.55), and monthly active users (Mean = 5181.62, Median = 0, SD = 14373.64). Apps on the Google Play Store also varied widely in their number of downloads (Mean = 249.09, Median = 0, SD = 1211.94), daily active users (Mean = 799.38, Median = 0, SD = 4054.71), monthly active users (Mean = 3048.16, Median = 0, SD = 15421.26). On the Apple App Store, the 5 most downloaded apps were responsible for 71% of downloads, 73% of daily active users, and 73% of monthly active users; on the Google Play Store, the 5 most downloaded apps were responsible for 96% of downloads, 99% of daily users. App store ratings were not associated with downloads, daily active users, and monthly active users ( $p > 0.37$ ). In the final paper, associations between MARS ratings, downloads, and active users will also be presented; all other analyses have been completed.

**Discussion:** We found that the top few apps were responsible for the majority of downloads, daily active users, and monthly active users, and most apps had zero active users. Our findings suggest that only the most engaging suicide apps are able to successfully acquire and retain users, implying that even interventions with “average” levels of engagement may not disseminate well. Surprisingly, app store star ratings were not associated with downloads or active users, suggesting that star ratings may not be a proxy for app engagement. Future research could examine if mobile app rating tools (see Torous et al., 2017) predict downloads and active users as one way to assess the external validity of these tools. Furthermore, our findings also have implications for reviews of digital suicide interventions. Reviews that systematically search the app store (e.g., Larsen, Nicholas, & Christensen, 2016) are useful to characterize what is available on the app store, but information on downloads and active users may be necessary to draw valid conclusions about the kinds of apps that individuals are actually using.

Given that a minority of apps is responsible for a majority of app traffic, future reviews of suicide apps may wish to focus on apps with high usage rates. Future research could examine whether these patterns exist in mHealth apps that target other behavioral health conditions. If so, understanding the characteristics that distinguish the most engaging digital health interventions from average digital health interventions may be a priority for mHealth researchers.

*Akash Wasil (Harvard University)*

## Independence D

### The Candid Body-Worn Camera? How Officer Body-Worn Cameras Inform (and Misinform) Our Understanding of Police Encounters

#### SYMPOSIUM

Police departments across the country have adopted body-worn cameras at a rapid pace. These recording devices are thought to hold officer and citizen alike accountable, providing a record of what actually transpires in police-community encounters. The four talks in this symposium address the promise and limitations of this footage in understanding police encounters: they can mislead as well as illuminate.

First, Yael Granot demonstrates how visual attention to evidence—as measured by eye-tracking technology—can attenuate or exaggerate bias against out-group members in the legal system. Across studies, participants watched a fight between a social out-group member and in-group member. Results showed that the more that people selectively attended to an out-group actor, the greater the discrepancy in decisions among differently identified individuals. Subtle visual instructions can also attenuate group-based bias punishment decisions, however: instructing viewers to watch these scenes relationally reduces racial disparities in punishment.

Kristyn Jones next describes how people bring their own biases to police footage. People may learn information from a police report, for example but then misremember it as being present in footage. Jones and colleagues test this possibility by presenting participants with BWC footage of an officer-involved shooting paired with a police report containing information that is not shown in the footage. Participants tend to “fill in” ambiguous parts of BWC footage with knowledge from other sources, and incorrectly remember unrepresented information based on their prior beliefs.

Broderick Taylor sheds light on a systematic bias introduced from body camera footage: that of visual perspective. Eight experiments reveal that body camera recordings result in lower observer judgments of intentionality than dash camera recordings of both scripted videos and naturalistic police recordings. This effect is due, in part, to variation in the visual salience of the focal actor: the body cam wearer is typically less visually salient when depicted in body versus dash cam video, which corresponds with lower observer intentionality judgments.

In the final talk, Nicholas Camp articulates a way in which these recordings can shed light on the everyday encounters between police and citizen. Using a combination of social psychological and computational analyses, Camp and colleagues use body camera footage to examine racial disparities in police officers’ language and test whether police officers communicate more respect towards White versus Black drivers. Three behavioral studies demonstrate the robustness of these disparities across participants and samples of officer communication. Using computational modeling, the authors use these ratings to uncover the linguistic correlates of respect and apply them to an entire month of officer speech: over 30,000 utterances.

Together, this symposium offers a nuanced and cutting-edge look at a technology that is rapidly changing law enforcement. By developing, testing, and applying psychological theory to body-camera footage, the talks here offer concrete suggestions to practitioners and raise additional questions for research.

*Chair: Nicholas Camp (Stanford University)*

#### Blue Words Matter: Using Body Camera Footage to Examine Racial Disparities in Officer Communication

The adoption of officer-worn body cameras has been promoted for accountability and evidentiary purposes. This focus on adjudicating officer responsibility in specific cases, however, overlooks the thousands of everyday interactions between police officers and citizens these cameras capture. Research from procedural justice demonstrates that the interpersonal respect officers communicate in these common encounters has consequences for citizens’ support and cooperation with law enforcement (Tyler & Huo, 2001). Combining both social psychological and computational analyses, we use body

camera footage to examine racial disparities in these contacts, focusing on one of most important tools at police officers' disposal: their words. In doing so, we could test whether police officers communicate more respect towards White versus Black drivers. In three behavioral studies, participants rated "thin slices" of officer communication: brief exchanges between officer and citizen, transcribed from camera footage of routine stops of Black and White drivers. Community and lab participants, who were blind to the race of the officer's interlocutor, rated over 800 of these utterances on several dimensions related to respect: the officer's respectfulness, politeness, impartiality, friendliness, and formality, which were combined into a single index of officer respect. Mixed-effects analyses found that officers communicated less respectfully with Black citizens, even controlling for relevant features at the stop, officer, and participant rater levels. These results replicated across participant populations and samples of utterances. We subsequently developed a computational model to predict participants' respect ratings using theoretically-derived linguistic features of respect, such as apologies and honorific titles. By correlating these features with participants' judgments, we were able to obtain the weights which best predicted officer respect, then use these weights to compute the respect for an entire month of traffic stops: over 30,000 utterances. Model results replicated behavioral findings, and also allowed us to map the trajectory of police-community interactions over time. Racial disparities emerged from officers' first utterances towards drivers, and increased over the duration of the stop. Together, these findings uncover racial disparities in everyday interactions which could erode trust between the police and communities of color. Further, they illustrate the utility of body-worn camera footage as data on the thousands of contacts between the police and community, and not simply evidence in a handful of them.

*Nicholas Camp (Stanford University)*

### **Body Camera Footage Leads to Lower Judgments of Intent Than Dash Camera Footage**

Police departments use body-worn cameras (body cams) and dashboard cameras (dash cams) to monitor the activity of police officers in the field. Video from these cameras informs review of police conduct in disputed circumstances, often with the goal of determining an officer's intent. Eight experiments (N = 2,119) reveal that body cam video of an incident results in lower observer judgments of intentionality than dash cam video of the same incident, an effect documented with both scripted videos and real police videos. This effect was due, in part, to variation in the visual salience of the focal actor: the body cam wearer is typically less visually salient when depicted in body versus dash cam video, which corresponds with lower observer intentionality judgments. In showing how visual salience of the focal actor may introduce unique effects on observer judgment, this research establishes an empirical platform that may inform public policy regarding surveillance of police conduct.

*Broderick Turner (Kellogg School of Management)*

### **Seen of the Crime: How Eye-Tracking to Video Evidence Can Predict Intergroup Legal Bias**

People have incredible faith in their visual experience, even in the often-uncertain context of legal decisions. Research suggests that people find video evidence to be uniquely credible and persuasive

compared to other forms of evidence. However, less is known about how the manner in which people watch such materials affects their evaluations and legal decisions. This talk explores how the manner in which people watch video evidence interacts with individual differences—in this case group identities—to predict legal judgments. In the first part of the talk, I demonstrate how visual attention to evidence—as measured by eye-tracking technology—can actually exaggerate bias against out-group members in the legal system. Across studies, participants watched a fight between a social out-group member and in-group member; videos of both police-civilian altercations as well as a fight between novel groups created in the lab were used. Results showed that the more that people selectively attended to an out-group actor, the greater the discrepancy in decisions among differently identified individuals. Only among individuals who attended frequently to the out-group actor did the expected pattern of out-group bias in legal punishment emerge. Thus, more narrowed attention to a social out-group member increases the influence of one's group identification on punishment decisions. In the second part of the talk, I show that subtle visual instructions can be used to attenuate group-based bias punishment decisions. White participants were presented with video evidence of an altercation between a White and Black actor, and were either given instructions to watch the video naturally or received our intervention instructions to watch the video relationally, equally attending to all aspects of the scene. White participants trained to watch a cross-race fight relationally assigned the Black actor significantly less jail and probation time than participants who watched naturally. This effect was qualified by a significant interaction between attention instructions and group identification: instructions to watch the scene relationally particularly reduced the punishment of participants strongly identified with their White in-group—those most likely to evince out-group bias in the control condition. Together, these studies elaborate the process by which people can come to different conclusions about the same information. In the context of a legal system where minority offenders are disproportionately likely to face harsher punishment, and video evidence is becoming increasingly ubiquitous, understanding the connection between technology and bias is critical. I discuss the implications of these findings for policy surrounding the use of video in the courtroom as well as for future avenues of interdisciplinary research.

*Yael Granot (Loyola University Chicago)*

### **Not in Sight: Source Monitoring Errors for Ambiguous Body-Worn Camera Footage**

Although police body-worn camera (BWC) footage is heralded as an objective record of critical incidents, research shows that people bring their own biases to police footage. Jones, Crozier and Strange (2017) found that a) people perceive BWC footage in part based on their pre-existing beliefs and b) that the events depicted in BWC footage can be distorted by an inaccurate police report describing the footage. They suggested source monitoring as a causal mechanism; that is people may learn information from a police report, but then misremember it as being seen in the BWC footage. Here, we tested that possibility by presenting participants with BWC footage of an officer-involved shooting paired with a police report containing information that is not shown in the footage. Additionally, we tested the propensity for participants to "fill in" ambiguous parts of BWC footage, and whether they would incorrectly remember unrepresented



information. Study 1 conformed to a 2(report: present, absent) x 2(footage: full, partial) factorial design. 528 participants from MTurk completed the survey. Half of the participants read a police report that described an officer-involved shooting—including describing the suspect having a gun that was not in the footage. Then, all participants watched BWC footage of the event; Half saw the entire footage; the remainder watched a version of the footage with 1s sections of the film removed at random points to create ambiguity about what was occurring. Finally, participants answered questions about their memory for and perception of the film. Specifically, whether they saw particular details in the footage and made ratings of whether they believed those details occurred in the event regardless of whether they saw it in the footage, -3 (very confident they did not see it in the video, or it did not happen) to +3 (very confident they did see it in the video, or that it did happen) with no midpoint. Participants also answered questions about their attitudes toward the officer's use of force, the extent to which the officer and civilian were responsible for the encounter, and completed the identification with police scale. We ran two MANOVAS—one for belief that participants saw the detail in the video, and one for the belief that the detail happened, with Report (present/absent) and Video (full/partial) as the IVs, Identification with Police score as a covariate, and responses to the four key items as the DVs. Overall, we found Identification with Police was a significant covariate for three of the four key items ( $p$ 's = [.001,.037], such that people with higher scores were more confident that they saw the key item in the film, and that it did indeed happen. Further, participants who read the report were more confident that the suspect reached for a gun after being shot ( $F=14.65$ ,  $p<.001$ ,  $\eta^2=.027$ ), that they saw the civilian put a gun in his waistband ( $F=6.02$ ,  $p=.014$ ,  $\eta^2=.011$ ), and that they saw the civilian hold a gun in the vehicle ( $F=7.88$ ,  $p=.005$ ,  $\eta^2=.015$ ) than participants who did not read the report. Surprisingly, whether the video was full or partial did not affect participants' ratings, and did not interact with the presence of the report ( $p$ 's = [.202, .783]. Overall, we found participants were more confident that the civilian had a gun (when he did not), and they saw the civilian have a gun in BWC footage, when they identified with police and when they read a report that described the civilian having a gun. The ambiguity of the video did not affect these false beliefs, demonstrating that people will not agree on what happened, even when the BWC footage is clear. To further this line of research, in Study 2, we are manipulating the plausibility of the misinformation items in the report to determine whether there are limits on people's source monitoring errors. Additionally, in Study 1, because we actually removed footage, participants could plausibly believe the missing content contained the key details. To overcome this limitation, we included blackout screens of the footage to create the illusion of ambiguity without excluding any critical information. Thus Study 2 was a 3 (report statements: plausible vs. less plausible vs. no report) x 2 (video: ambiguity vs. no ambiguity design) between subject design. Data collection is not yet complete, but we hypothesize participants in the ambiguous condition will make more source monitoring errors than those in the nonambiguous condition, that participants who are less likely to commit source monitoring for implausible statements, and that identification with police will once again predict the extent to which people endorse items as occurring, even if they didn't read about them. To summarize our data suggest that BWC footage is not a cure-all for problems between civilians and police. Additionally police and media policies for BWC footage are currently inappropriate given that seeing is not always believing.

*Kristyn Jones (John Jay College of Criminal Justice)*

## Independence FG

### Transforming Access to Mental Health Care for Rural and Underserved Populations

#### SYMPOSIUM

In the U.S., there is a shortage of mental health care providers, particularly in rural areas (Andrilla et al., 2018). The unequal distribution of providers puts individuals living in a rural area at a disadvantage for receiving mental health care. Having limited options can also lead to sizable costs as many people turn to emergency room care for treatment (Insel, 2008; Turner & Stanton, 2015). Telehealth has been used as a solution to this problem by increasing access to care to more areas. Our clinic, the Texas A&M Telehealth Counseling Clinic, serves 7 rural counties in central Texas while also training future practitioners in telehealth practices. In this session, we will look at the need for addressing this disparity and the potential telehealth has to help, discuss our model for a telehealth training clinic, and our clinic's outcomes.

The first presentation addresses in more depth the national need for addressing disparities and the potential for telehealth to reduce these disparities. Research shows that many individuals have a need for mental health care, but are unable to access care (Mental Health America, 2018; SAMHSA, 2018). Rural America is especially affected by this as most providers live in urban cities. Telehealth can help to address these issues related to access to mental health care by using telecommunication to provide health care across distances (Nickelson, 1998). This presentation will outline and further explain the national need for mental health care access and how telehealth is and can continue to help meet this need.

The second presentation proposes a model for a telehealth counseling training clinic based on the Texas A&M Telehealth Counseling Clinic. The clinic is built on 3 foundational goals: 1) telebehavioral health service delivery, 2) policy relevant research and evaluation, and 3) interdisciplinary training. This combination allows for building a future of telepsychologists while also being able to offer free counseling services to clients who often have no access to other care and have no insurance. We utilize a combination of a hub and spoke model and in home model. The hub and spoke model includes having partnerships with local resource centers or clinics where clients go to see their counselor over videoconferencing. In having both models available, we are better able to provide equitable access for the underserved. The clinic is preparing the next generation of psychologists to be prepared for telepractice and the reality that healthcare will be provided in teams and psychologists should be prepared to work across disciplines to solve healthcare problems.

The third presentation describes how our clinic integrated model within two primary care settings. This integration is highly valuable as the presence of tele-psychologists in rural health care eases the burden on primary care physicians, who are often first line of care in rural communities, as well as increase patients access to high-quality mental health care. Many individuals in rural areas suffer from chronic health conditions due to lack of resources, and integration of telecounseling into these patients' treatment plans may work to increase health outcomes and reduce risk for comorbid mental health concerns. The use of telepsychology allows for clinicians to provide

these services at multiple rural sites, reducing the burden on the client and the referring physicians.

The final presentation looks at our clinic's outcome data in regards to number of sessions needed for progress and views about their experience at our clinic. Although all of this data is not generalizable, it is beneficial to examine and understand so to inspire future research. Overall, we can learn from the data that our clients likely would not have access to mental health services without our clinic and that they had a positive experience using telehealth. Additionally, we can see that based on outcome measures like the PHQ-9 and the CORE-B, clients may only need anywhere from 5 to 11 sessions to improve, showing that brief therapy can be beneficial. This allows us to limit the amount of sessions offered and serve more people. These results also provide additional evidence for telehealth's effectiveness as a service delivery modality.

Due to the growing need in access to mental health care, we propose a call to action for additional research and practice of telehealth. This session will encourage attendees to expand the use of technology in their practice by describing a telepsychology research, services, and training model. By combining technology and mental health care, we can have an impact on individuals and systems and reduce disparities in access and availability of services.

*Chair: Carly McCord (Texas A&M University)*

#### **Addressing Disparities in Access to Mental Health Care Using Telehealth**

According to the most recent available data from the National Survey on Drug Use and Health, almost 19% of adult Americans have any mental illness (AMI), and approximately 4.5% of adult Americans have a serious mental illness (SMI; SAMHSA, 2018). Among individuals with an AMI, only 42.6% received mental health services, and approximately 66.7% of adults with an SMI received mental health services. Approximately 13.5 million American adults had a perceived unmet need for mental health care, and one out of five adults with a mental illness report that they are not able to receive the treatment they would like (Mental Health America, 2018; SAMHSA, 2018). These disparities exist despite a majority of Americans endorsing mental health as equally important as physical health (America's Mental Health, 2018). However, many Americans do not know where to go to receive treatment, or are afraid of judgement from others if they were to seek out treatment (America's Mental Health, 2018). The largest barrier to receiving mental health treatment is a lack of access. This includes a lack of accessible providers and not being able to afford treatment. This lack of accessible and affordable providers is partly explained by the shortage of mental health care providers in the U.S., especially in rural areas (Andrilla et al., 2018). The Health Resource and Service Administration designates areas in the US that are lacking in readily accessible mental health care options as "Mental Health Professional Shortage Areas (MHPSAs)" (U.S. Department of Human Services, 2018). In the U.S., there are approximately 5,000 MHPSAs, which is home to over 115 million Americans. The unequal distribution of providers puts individuals living in a rural area at a disadvantage for receiving mental health care. Having limited options can also lead to sizable costs as many people turn to emergency room care for treatment (Insel, 2008; Turner & Stanton, 2015). Other societal costs of mental illness include early mortality, lost productivity and earnings, and increased disability. Telehealth can help to address these issues related to access to

mental health care by using telecommunication (i.e. phone and videoconferencing) to provide health care across distances (Nickelson, 1998). By utilizing technology, telehealth can reach more individuals who are in need of services, especially those in MHPSAs. Research supports that telehealth improves access to care for individuals, especially those in rural areas or those without access to professional mental health resources, as well as to patients who may require specialized care, such as children, adolescents, or the elderly (Bashshur, Shannon, Bashshur, & Yellowless, 2016). Telehealth has also been found to be efficient and cost effective, especially when utilized as a collaborative effort between healthcare providers (e.g., primary care physicians) and with larger volumes of patients (Bashshur et al., 2016). Telehealth has also been shown to improve mental health care in primary care settings, and can improve patient compliance with treatment and medication adherence. It has also been shown to effectively treat a variety of mental disorders, and improve outcomes for individuals who have comorbid medical conditions (e.g., diabetes, hypertension, etc.). Overall, there is a demonstrated need for better access to mental healthcare services in the United States, especially in rural areas, and telehealth is an effective tool to address this need. In fact, the American Telemedicine Association's Guidelines for Video-Based Online Mental Health Services (2013) reported that "to date, no studies have identified a patient subgroup that does not benefit from, or is harmed by mental healthcare provided through remote video conferencing." Telehealth is a beneficial way to overcome access barriers to treatment for the millions of individuals in the US with AMIs, SMIs, and other circumstances in which they may require mental health care. This presentation will outline and further explain the national need for mental health care access and how telehealth is and can continue to help meet this need.

*Laylee Jackson (Texas A&M University)*

#### **Establishing and Developing a Telepsychology Clinic in Texas**

The purpose of this section of the symposium is to introduce a telepsychology service delivery and training model to stimulate the development of other service models nationally. The mission of the Telehealth Counseling Clinic is three fold. The clinic strives to lead in 1) telebehavioral health service delivery, 2) policy relevant research and evaluation, and 3) interdisciplinary training. We are committed to innovative solutions for the evolving health needs of our constituents. The clinic started as an interdisciplinary collaboration between the School of Public Health and the College of Education and Human Development Counseling Psychology program at Texas A&M University. The Center for Community Health Development at the School of Public Health routinely conducted community health assessments of the region and found little to no access to mental health services, lack of transportation, and poorer than average mental and physical health statuses. A Health Resources and Services Administration grant funded the first pilot project in 2009 to link counseling psychology doctoral students to the community with the greatest need in the region via videoconferencing. Since then, additional HRSA funds and the Texas 1115 Waiver have allowed the clinic to grow to serve 7 counties via 9 access points. We have provided over 10,000 sessions to about 1,000 individuals. Services are provided via videoconference and telephone to individuals aged 13 and up in English, Spanish, and Mandarin. We utilize a combination of a hub and spoke model and in home model. This is both practically and

clinically needed. Some clients lack the technology or broadband to connect from home or have clinical presentations such as depression that benefit from behavioral activation and getting out of the home or tendencies towards self-harm that prevent them from keeping themselves safe during triggering sessions. These individuals are best served by visiting an access point in a hub and spoke model. Other individuals lack transportation or childcare or have clinical presentations such as agoraphobia and would have previously gone without treatment if required to visit a site, but are good candidates for in-home treatment. In having both models available, we are better able to provide equitable access for the underserved. Our videoconferencing platform allows for the remote collection of intake paperwork and routine outcome monitoring and is specifically designed for rural areas by requiring a very low bandwidth of .25mbps to maintain a connection. To be successful as a clinician in a rural area, one must be a generalist and ready to treat a wide variety of presenting concerns. We are constantly faced with the ethical dilemma of working within the boundaries of our competence balanced with not doing harm by referring to a client to treatment that in reality is inaccessible to them. The most common presenting concerns seen at the clinic are depression, anxiety disorders, bipolar disorder, substance abuse, and trauma. The clinic is preparing the next generation of psychologists to be prepared for telepractice and the reality that healthcare will be provided in teams and psychologists should be prepared to work across disciplines to solve healthcare problems. The consumer market is readily adopting telehealth, but our academic institutions are not progressing at the same rate. The clinic provides practicum training to psychology doctoral students and continuing education to licensed providers. Additionally, the clinic provides students from a variety of disciplines including psychology, engineering, health education, architecture, and public health to apply what they are learning in the classroom to a clinic setting to positively impact the communities we serve. The clinic has trained over 60 doctoral psychology students, of which some now have careers in telepsychology in the Department of Defense, Veterans Administration, and private practice. The symposium will describe the community capacity building model that was followed in establishing and growing the clinical services and provide specific examples of how students from different disciplines are being trained.

*Carly McCord (Texas A&M University)*

### **Integrating Telepsychology in Rural Primary Care Settings**

The purpose of this presentation is to discuss the use of telepsychology within a rural primary care setting. Phone and video conferencing services allow the Telehealth Counseling Clinic to reach individuals from 7 distinct counties in the Brazos Valley. Most of these individuals would not otherwise have access to mental health care. In collaboration with Texas A&M Physicians (TAMP) and Health for All (HFA), two primary care clinics in Brazos County, the TCC has expanded services through the Texas 1115A Medicaid Waiver Delivery System Reform Incentive Payment (DSRIP) program and worked to ease the burden on primary care physicians (PCPs) in rural Texas. In addition to the severe shortage of licensed mental health care providers, such as psychologists and psychiatrists, Texas is 47th in the number of primary care physicians per 100,000 population. Nearly 60% of these physicians reside in only 5 counties, leaving the rest of Texas largely underserved. Few qualified health professionals are available to meet Texas residents' physical and mental health

needs in most of the state. Further, Texas is 48th in the U.S. for health funds distribution (North Texas Regional Extension Center, 2015). This places a significant strain on primary care physicians (PCPs), who are often the first line of care for rural patients. In rural areas, patients are increasingly looking to PCPs for diagnosis and treatment of mental health concerns. Many individuals suffering from psychological conditions, such as posttraumatic stress disorder (PTSD), first seek help from their PCP. Yet due to lack of resources and training, many conditions go undiagnosed and untreated. For example, studies estimate that of those who meet criteria for PTSD, only 2% receive a diagnosis and treatment referral for their symptoms from their PCP (Taubman-Ben-Ari, Rabinowitz, Feldman, & Vaturi, 2001). Receiving psychological services may also improve physical health outcomes. Patients who experience chronic health conditions (e.g., diabetes) may be at higher risk of experiencing depression or other mental health problems. These concerns may also negatively impact their disease management. Studies have found mental health services to be beneficial for patients with diabetes, helping them to have better glycemic control (Rubin, 2000). By integrating telehealth services (i.e., TCC services) into primary care settings (i.e., TAMP), patients are better able to manage their comorbid conditions. Through the DSRIP program, the TCC has been able to focus on counseling interventions for patients of TAMP and HFA. TAMP PCPs refer clients for telecounseling, and to facilitate collaborative care, TCC clinicians utilize a shared system to document notes. Client outcome data, including A1C levels (for diabetes clients) and PHQ-9 scores, are collected at regular intervals to gauge client progress. Thus far, the clinic has provided a total of 456 counseling sessions to 67 unique TAMP patients, for an average of 6.8 sessions per client (SD = 8.2). All TAMP clients are seen via phone and video conferencing modalities, as patients live in multiple counties throughout the Brazos Valley. Clients vary in ethnicity (White:  $n = 33$ , 49.3%; Latinx:  $n = 18$ , 26.9%; African American:  $n = 13$ , 19.4%; more than one/unknown/other:  $n = 3$ ; 4.5%). A majority of clients are female ( $n = 58$ ; 86.6%), and they range in ages ( $M = 43.59$ ,  $SD = 13.87$ ). Mental health diagnoses have included depression (49.5%), anxiety (22.6%), PTSD/adjustment disorder (7.5%), substance abuse disorder (2.2%), and other serious mental illnesses (18.3%). Comorbid chronic health conditions were present in 39.39% of TAMP clients (e.g., diabetes and hypertension). At HFA the TCC utilizes a dual approach, with one practicum counselor seeing patients on location and another counselor seeing patients via video conferencing. This allows the TCC to see twice the number of clients and take warm hand-offs from PCPs who would like their patient to be evaluated for counseling. At times, the TCC has participated in HFA physician huddles and provided consultation on cases regarding mental health concerns.

*Mariah Stickley (Texas A&M University)*

### **Measuring the Impact of Telehealth Service Delivery**

**Introduction/Statement:** There is preliminary evidence showing the effectiveness of telepsychology services across the board in terms of client outcomes; so far, the research is encouraging (Hilty et al., 2013). While promising, there remains a need for additional research on the effectiveness of telepsychology services, specifically televideo services. One psychotherapy outcome model by Jacobson and Truax (1991) consists of two major criteria: 1) after receiving therapy, clients must cross over from a clinical range into a non-clinical range, and 2) client outcomes must reliably improve between pre-therapy

and post-therapy based on a reliable change index (Jacobson & Truax, 1991). They outline four outcomes including clinically significant change (CSC; improvement that also moves the client into a non-clinical group), reliable improvement (RI), reliable deterioration (RD), and no reliable change. Overall improvement can be conceptualized as looking at individuals who have achieved reliable improvement (RI) regardless of their starting or ending point. This does not have the stricter requirement of CSC that requires movement from the clinical to normal sample by the end of treatment. Because of this leniency, it is the best representation of overall counseling improvement. The primary question of interest here is, How many telehealth counseling sessions did clients receive on average until they reached clinically significant change or reliable improvement on the PHQ-9 or the CORE-B (total and subscale scores)?

**Procedures:** Participants were clients receiving services from the Telehealth Counseling Clinic (TCC) located in College Station, TX. Clients must first complete a phone screening during which telepsychology services are described, demographic information is collected, and contact information is obtained from the client. The starting sample consisted of 413 clients, but only 155 clients attended at least two sessions and could be further analyzed. The Patient Health Questionnaire (PHQ; Kroenke, Spitzer, & Williams, 2001) and CORE-B Short Form (CORE System Group, 1998) are administered at each intake appointment. Clients then receive appropriate weekly counseling through videoconferencing or telephone by TCC counselors. On an ongoing basis, clients completed the CORE-B Short Form weekly and the PHQ-9 every two weeks at the time of their counseling sessions. Satisfaction surveys were administered after every third session and after therapy termination. Measures: CORE-B. An 18-item self-report measure of functioning in multiple areas: well-being, problems, life functioning, and risk to self and others. An individual's score is calculated in each of the four subsections and a total mean with scores ranging from 0-4. PHQ-9. Clients were administered the PHQ-9 depression screener every two sessions. This is a quick, depression screener measure based on DSM-IV diagnostic criteria. A total score ranges between 0 and 27 and is calculated by summing the 9 items. Items are on a 4-point Likert-scale ranging from 0 (Not at all) to 3 (Nearly every day). Satisfaction Surveys. Satisfaction surveys aimed at assessing client satisfaction of therapy services, satisfaction with the technology modality employed, and access to quality mental health care. 203 surveys have been returned and analyzed. Analyses: Demographic characteristics of the sample will be provided. Number of sessions needed to achieve improvement from therapy was estimated using Kaplan-Meier survival curves; specifically, the median survival time provides an overall estimate. Descriptive statistics were also utilized for the satisfaction survey results. Results: Clients achieved either CSC or RI on the PHQ after an average of 11 sessions (SE = 2.39, 95% CI [7, 19]). For the CORE-B total score, clients improved in four sessions (SE = .54, 95% CI [3, 6]), and on the other subscales, clients improved in four to seven sessions. Clients reported that the quality of services received was high (96% Strongly Agree/Agree) and comparable to in-person, face-to-face counseling (90% Strongly Agree/Agree). Clients also reported convenience in receiving telehealth services compared to traveling to a similar provider (85% Strongly Agree/Agree). Last, many of our residents (77% Strongly Agree/Agree) would have gone without services without the provided telehealth services. Practical Implications: Outcomes showed positive results in multiple areas for our clients. Results from this study generally support other research showing that 11 to

14 sessions of psychotherapy lead to clinically significant change or reliable improvement on outcome measures, and additionally, there is much client growth in the first few sessions. One reason why the CORE-B scale/subscales may have shown earlier CSC/RI events is because of its multifaceted nature. These results provide additional evidence for telehealth's effectiveness as a service delivery modality. Information on number of sessions and other "data-driven" sources are also utilized at the TCC in its clinic policy decisions. Further, the TCC also uses these types of analyses as "hard evidence" to funding agencies for how they support their community, reduce usage of other vital community resources, such as emergency departments, help reduce barriers to quality mental health care in this area, and provide information on continual improvements and areas needing additional attention/funding in the Brazos Valley.

*Katie Console (Texas A&M University)*

---

### Lafayette Park

## Reflections on Cyberlearning: Exploring Tensions in the Co-Evolution of Learning Technologies with Advances in Learning Theories and Methods

### PANEL

Research in "cyberlearning" investigates technologies on the horizon, along with the advances in theory and methods that will be needed to create powerful learning opportunities. This session builds on discussions at a conference earlier in the week, "Cyberlearning 19: Exploring Contradictions in Achieving Equitable Futures." It will invite discussion of the tensions and challenges that are essential to investigating the future of learning with technology. Three presenters will each share an example of their own research, in order to ground the discussion in the work that cyberlearning researchers do.

Cynthia D'Angelo will communicate some of the core challenges that arise in her projects that involve using speech data in classroom settings. The goal in these projects to use speech data to better understand learning, teaching, and collaboration. The projects typically analyze the speech activity (e.g., the amount of talk, the amount of overlapping talk, distribution of talk across speakers) of teachers and students. This activity can provide information about how well small groups are collaborating with each other or how well a teacher is facilitating a classroom discussion. Making this invisible data visible for teachers and providing frameworks to help teachers act on this data is an important ongoing aspect of the work. Current tensions in this work are issues surrounding data privacy and making sure students are aware of how, when, and why their data is being collected and how it will and will not be used.

Tammy Clegg will discuss the Science Everywhere project, a five-year design based research project focused on designing social media tools and community displays. The design intention is to use social media and public displays to promote science learning across contexts for youth (K-12) and families in low socio-economic communities — contrasting with a typical focus of learning science only within the walls of particular classrooms, largely as a private activity. Within this context, the Science Everywhere research team focuses

on understanding youth and community members' (e.g., parents, teachers, community volunteers) science disposition development as they interact with Science Everywhere technology and associated learning experiences within two neighborhood settings. Clegg will highlight tensions the research team has identified as inherent in expanding community stakeholders' learning roles (e.g., parents as caregivers and learning partners) in feasible ways. Additionally, she will discuss tensions realized as youth and families learn in new neighborhood spaces (e.g., churches).

James Lester will discuss a project titled "Augmented Cognition for Teaching: Transforming Teacher Work with Intelligent Cognitive Assistants." This project focuses on the future of work for teachers and asks: how can we leverage AI to improve not only teachers' performance but also teacher quality of work-life? The project centers on the design, development, and evaluation of the Intelligent Augmented Cognition for Teaching (I-ACT) framework for intelligent cognitive assistants for teachers. With a focus on assisting K-12 STEM teachers in technology-rich inquiry teaching that supports collaborative, problem-based STEM learning, I-ACT cognitive assistants provide teachers with (1) prospective pedagogical guidance (preparation support preceding classroom teaching), (2) concurrent pedagogical guidance (real-time support during classroom teaching), and (3) retrospective pedagogical guidance (reflection support within a community of practice following classroom teaching). A key tension in the work is how to create cognitive assistants for teachers that will leverage multimodal classroom data streams that serve a robust support function but are not perceived as "evaluative" by teachers — or as a surveillance mechanism.

*Chair: Judi Fusco (Digital Promise Global)*

*Cynthia D'Angelo (University of Illinois)*

*James Lester (North Carolina State University)*

*Tamara Clegg (University of Maryland)*

*Discussant: H. Chad Lane (University of Illinois)*

4:30–5:30 P.M.

**Independence Ballroom**

## Closing Panel

---

### Closing Session: Reflections, Connections, and Future Directions

#### PANEL

Come together to wrap up APA's 2019 Technology, Mind, and Society conference with a facilitated panel that will explore key takeaways from the conference, emerging issues, opportunities, and challenges, future directions for research and practice, and maybe even a sneak peek at what's in store for TMS 2020!

*Chair: David W. Ballard (American Psychological Association)*

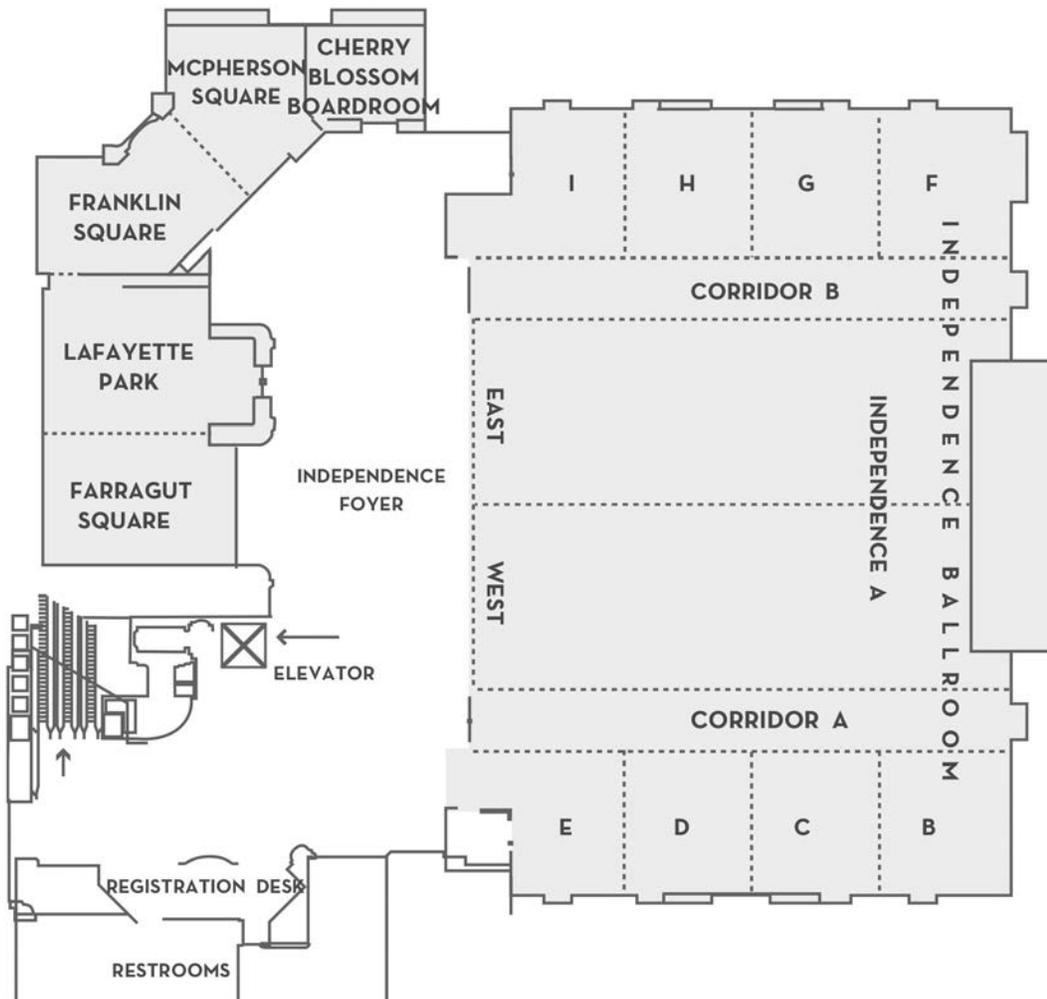
*Tara Behrend (George Washington University)*

*H. Chad Lane (University of Illinois, Urbana-Champaign)*

*Russell Shilling (American Psychological Association)*

## GRAND HYATT WASHINGTON

### Independence Level



#### Registration, Exhibits, Networking

Independence Foyer

#### Keynotes, Opening Session, Closing Panel

Independence Ballroom

#### Concurrent Sessions

Lafayette Park

Farragut Square

Independence BC

Independence D

Independence E

Independence FG

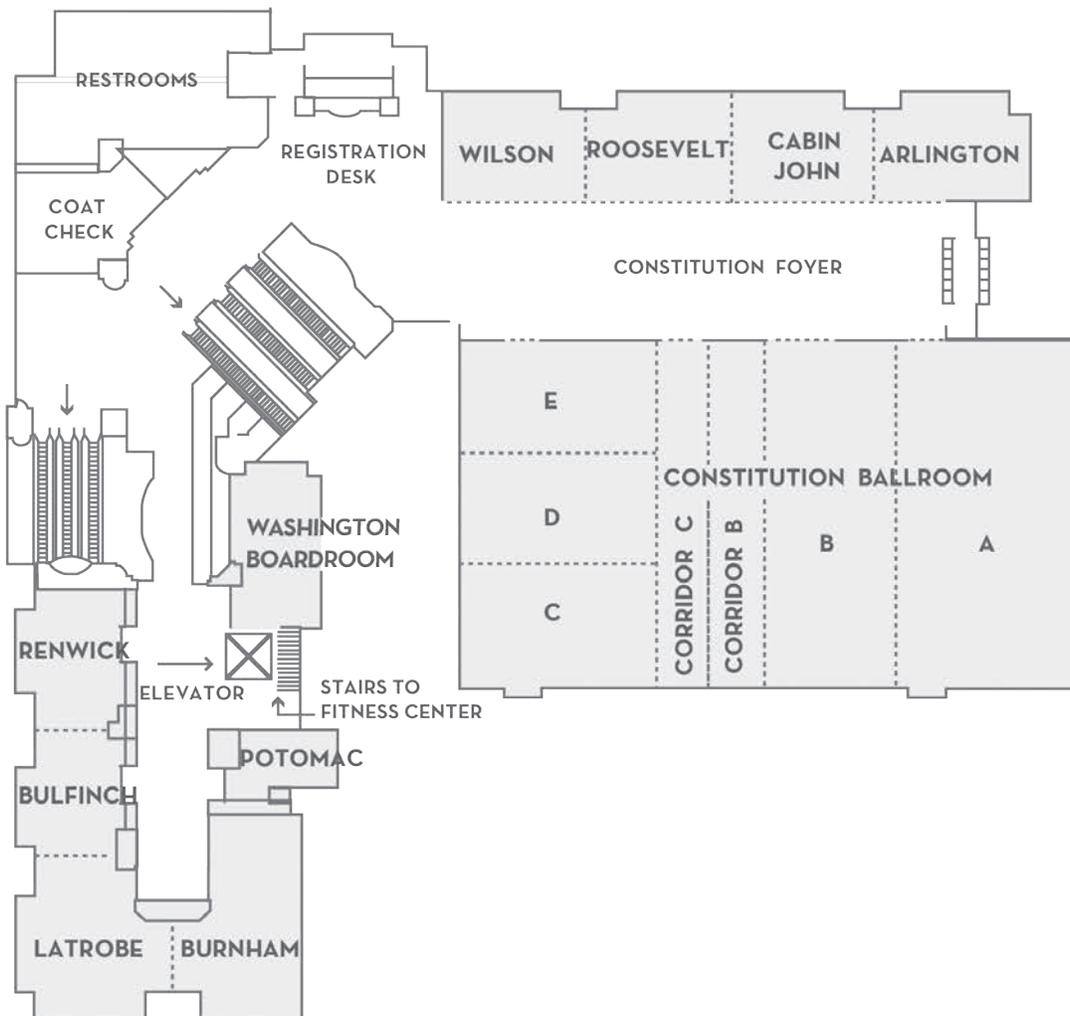
Independence HI

#### Speaker-ready

Franklin Square

# GRAND HYATT WASHINGTON

## Constitution Level



### Poster Sessions and Continental Breakfast, Welcome Reception

Constitution Ballroom

### Interview Room

Renwick

### Press Room

Potomac



**HIGH PERFORMANCE IT EXPERTS SINCE 2001**  
**CLOUD • STORAGE • SECURITY • NETWORKING**

**REGISTER TODAY!**

**THE DMV  
CYBER SECURITY SUMMIT**

**Friday, October 11  
9am - 1pm  
MGM National Harbor**

***planbtech.net***

Join the Mailing List  
for Updates on TMS 2020.

[TMS.APA.ORG](https://TMS.APA.ORG)

