46th Annual Meeting of the Society for Computers in Psychology

Boston Sheraton hotel
November 17th, 2016
Welcome Message from the President

Welcome to the 46th Annual Meeting of the Society for Computers in Psychology (SCiP).

This year’s theme is Cognitive Technologies. I selected the theme to celebrate the fact that computational psychology has grown into a new era where it can make meaningful contributions not only to our understanding of cognition and behaviour, but also to the development of applied cognitively-inspired technologies. The recent work on computational approaches to semantic memory and the use of those models to conduct semantic text mining provides one excellent example of this emerging endeavor.

With this in mind, I have invited leading and emerging researchers to tell us about the work they are conducting to translate knowledge about basic psychological processes to develop cognitive technologies. Our Keynote Speaker, Art Graesser, requires no introduction and has long been at the vanguard of the cognitive technology effort. It’s a pleasure to have him present his work: a model for how to conduct applied cognitive science. Brendan Johns from the University at Buffalo is a more recent advocate and will discuss his efforts at semantic text mining. Dr. Evimaria Terzi from Boston University will discuss her work that uses computing to understand team intelligence and dynamics. Of course, those talks complement an already outstanding program of work that is pushing the boundaries of computational methods, experimental technique, and scientific knowledge.

I hope that you are as excited as me to meet with old friends and make some new ones. If you are new to the conference, please do not hesitate to ask me or members of the Steering Committee questions or to provide your thoughts on the conference.

Finally, I would like to remind you that all presenters, whether of spoken papers or posters, are invited to submit manuscripts based on the work they present to the Special Issue of Behavior Research Methods (see scip.ws for details).

Enjoy SCiP 2016 and your visit to Boston!

— Randy Jamieson
**SCiP Officers**

President: Randy Jamieson  
President Elect: Xiaowei Zhao  
Past President: Pete Dixon  
Secretary-Treasurer: Rick Dale

**Steering Committee**

Carol Connor, Arizona State University  
Ulf-Dietrich Reips, University of Konstanz  
Jennifer Trueblood, Vanderbilt University  
Matthew Crump, Brooklyn College of CUNY  
Stephanie Huette, The University of Memphis  
Jon Willits, University of California Riverside  
Scott Crossley, Georgia State University  
Chris Westbury, University of Alberta  
Todd Gureckis, New York University

---

**General information**

The Society for Computers in Psychology is a non-profit organization of researchers interested in applications of computers in psychology. Its primary purpose is to “increase and diffuse knowledge of the use of computers in psychological research.” Over the past several years the organization has set a special goal of aiding psychologists in using computational methods in their teaching and research. We have also encouraged consideration of the psychological aspects of hardware and software development and design. Membership is open to any person who has an academic degree and who is active in scientific applications of computers to psychological research.

**SCiP Time**

In order to allow people to visit presentations in different sessions we need to keep a tight schedule. For this reason, we will be timing each presentation according to SCiP time. SCiP time is determined by the clock displayed at the secretarial desk in the central hall.

**Presentation Guidelines**

Digital projectors will be available in each of the talk session rooms. Speakers should test their computers and presentation software during the break prior to their talk to ensure that there are no compatibility problems and that they will be able to begin their talk at the scheduled time without delays. Talks should be planned for a duration of 12 minutes, with an additional 3 minutes allowed for questions.

Poster presenters should set up their poster in the period from 12:15pm to 1:00pm and should be available at their poster from 1:00pm to 2:10pm. Posters should be taken down by 2:15pm so conference attendees have time to make their way to the President’s Symposium. Posters should be formatted to fit in a 6’ wide x 4’ tall. The poster session will be held on Thursday, November 17th, from 1:00 - 2:10pm. The locations of the posters are numbered as in the program.

**Presidential symposium**

Dr. Brendan Johns, University of Buffalo  
Dr. Evimaria Terzi, Boston University

**Keynote**

Dr. Arthur Graesser, The University of Memphis
POSTER SESSIONS WILL BE HELD IN BALLROOM A-C OF THE HYNES CONVENTION CENTER - LEVEL THREE

Posters in the Hynes Convention Centre
<table>
<thead>
<tr>
<th>Time</th>
<th>Fairfax A</th>
<th>Fairfax B</th>
<th>Berkeley</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:30 - 8:00</td>
<td>Registration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:00 - 9:00</td>
<td>Statistics with R Chair: David Vinson</td>
<td>Games and Web Chair: Stan Reimers</td>
<td></td>
</tr>
<tr>
<td>8:00 - 8:15</td>
<td>Using R to explore ERP data</td>
<td>Use of Warnings for Instructing Users</td>
<td>Aliping Xiong, Robert W. Proctor, Ninghui Li, &amp; Weining Yang</td>
</tr>
<tr>
<td>8:15 - 8:30</td>
<td>An R package to conduct LME based item-level analysis of performance in artificial grammar experiments</td>
<td>Human Trust in Automation in a Phishing Context</td>
<td>Jing Chen, Robert W. Proctor, &amp; Ninghui Li</td>
</tr>
<tr>
<td>8:30 - 8:45</td>
<td>Silky - A rich, graphical alternative to SPSS, which unifies graphical statistics software and the R command line.</td>
<td>Can playing video games produce flow experiences?</td>
<td>J. Scott Jordan &amp; Devin Gill</td>
</tr>
<tr>
<td>8:45 - 9:00</td>
<td>Using complicated n-gram smoothing algorithms efficiently in R</td>
<td>A gamified mobile web app for real-time in-lecture quizzes</td>
<td>Stan Reimers</td>
</tr>
<tr>
<td>9:00 - 9:15</td>
<td>Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:15 – 10:45</td>
<td>Computational semantics Chair: Chris Westbury</td>
<td>Eye tracking and classification Chair: Joseph Houpt</td>
<td>Internet symposium Chair: Ulf-Dietrich Reips</td>
</tr>
<tr>
<td>9:15 - 9:30</td>
<td>They Got Inside my Head: Using Knowledge Representations Derived from Text to Analyze Open-Ended Survey Responses</td>
<td>Gaze Path Segmentation with the Beta Process Autoregressive Hidden Markov Model</td>
<td>Joseph W. Houp &amp; Mary E. Frame</td>
</tr>
<tr>
<td>9:30 - 9:45</td>
<td>Automated computation of continuous semantic categories from unambiguous labelled exemplars</td>
<td>A minimal simulation model of foveated eye movements and visual search</td>
<td>Shane T. Mueller, Yin Yin Tan, Hannah North, &amp; Kelly Steelman</td>
</tr>
<tr>
<td>9:45 - 10:00</td>
<td>Automated Summary Assessment: Comparison of LSA Similarity with Varied References</td>
<td>ROC analysis of eye movements on complex images mediated by linguistic context</td>
<td>Stephanie Huette &amp; Jeffrey Vlaut</td>
</tr>
<tr>
<td>10:00 - 10:15</td>
<td>Cross-situational learning in a Zipfian environment</td>
<td>Using Best/Worst Scaling to Collect Semantic Judgments</td>
<td>Geoff Hollis</td>
</tr>
<tr>
<td>10:15 - 10:30</td>
<td>The Associative Boost in Semantic Priming: May It Rest in Peace</td>
<td>Vocal emotion classification based on self-organizing map algorithms</td>
<td>Xiaowei Zhao, Zhi-Yong Tao, &amp; Ya-Xian Fan</td>
</tr>
<tr>
<td>10:30 - 10:45</td>
<td>The Center for Data on the Mind: Helping cognitive scientists tackle the promise and challenges of big data to explore behavior and cognition</td>
<td>The QC-methodology: adapting the Q-methodology for analyzing Concepts through Categorical data</td>
<td>Margot Brard &amp; Sébastien Lê</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Svendroid: A generic smartphone app configurator for mobile assessment studies</td>
</tr>
<tr>
<td>Time</td>
<td>Fairfax A</td>
<td>Fairfax B</td>
<td>Berkeley</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>10:45 - 11:00</td>
<td>Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00 - 12:15</td>
<td>Experimental measurement Chair: Felix Henninger</td>
<td>New measures: Social and cognitive Chair: Nathan Evans</td>
<td>AutoTutor symposium Chair: Xiangen Hu</td>
</tr>
<tr>
<td>11:00 - 11:15</td>
<td>lab.js -- A Modular Foundation for Browser-Based Experiments Felix Henninger</td>
<td>Bayes factors for the Linear Ballistic Accumulator Model of Decision-Making Nathan J. Evans &amp; Scott D. Brown</td>
<td>Art History Danielle McNamara</td>
</tr>
<tr>
<td>11:15 - 11:30</td>
<td>Metronome LKM: An open source virtual keyboard driver to measure experiment software latencies Pablo Garaizar &amp; Miguel Vadillo</td>
<td>Group communication analysis: A computational-linguistic framework for exploring conversational roles in online multi-party communication Nia Dowell, Tristan Nixon, Zhiqiang Cai, &amp; Arthur Graesser</td>
<td>Impact of Pedagogical Agents' Conversational Formality on the Formality of Written Summaries Haiying Li &amp; Art Graesser</td>
</tr>
<tr>
<td>11:30 - 11:45</td>
<td>Capturing the Writing Process: Keystroke Logging in a Writing Tutor Laura K. Allen, Matthew Jacovina, Mihai Dascalu, Rod Roscoe, Kevin Kent, &amp; Danielle S. McNamara</td>
<td>Context dependent measures of Synchrony in Team Sports Maurici Lopez-Felip, Tehran J. Davis, &amp; Till D. Frank</td>
<td>Understanding sentence and text level factors that affect performance on the Reading Strategy Assessment Tool Joe Magliano, Danny Feller, Laura Allen, Keith Millis, &amp; Danielle McNamara</td>
</tr>
<tr>
<td>11:45 - 12:00</td>
<td>Mousetrap: Free, open-source, and cross-platform mouse-tracking and analysis plug-ins Pascal Kieslich &amp; Felix Henninger</td>
<td>Examining the impact of tutor dialogue moves on knowledge in a Wizard of Oz system Colin L. Widmer, Christopher K. Wolfe, Cecelia M. Favede, &amp; Richard Smith</td>
<td>Integrating the BRCA Gist Intelligent Discourse System into Routine Workflow in Clinical Settings Christopher R. Wolfe &amp; Elyse E. Lower</td>
</tr>
<tr>
<td>12:00 - 12:15</td>
<td></td>
<td></td>
<td>AutoTutor Question Answering</td>
</tr>
<tr>
<td>12:15 - 1:00</td>
<td>Lunch and poster setup</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:00 - 2:15</td>
<td>Poster session</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Ballrooms, Level 3 in Hynes Convention Centre)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:30 - 3:45</td>
<td>Presidential Symposium Brendan Johns Evimaria Terzi</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Fairfax A + Fairfax B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:45 - 4:00</td>
<td>Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4:00 - 5:00</td>
<td>Keynote Presentation Arthur Graesser</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Fairfax A + Fairfax B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:00 - 5:30</td>
<td>Business meeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Fairfax A + Fairfax B)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Posters**  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>***** Ballrooms A+B+C (Level 3) in Hynes Convention Centre *****</td>
<td></td>
</tr>
<tr>
<td><strong>1</strong></td>
<td>Given Names and Systemic Bias Replication through a Semantic Model</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>Exploratory Text Analysis with R and Shiny</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td>WhatsOb: An Android app for note-taking during behavioral observations</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>Evaluating Equating Methods for a Manual and Computerized Card Sorting Task</td>
</tr>
<tr>
<td><strong>5</strong></td>
<td>Attention Motivation through Gamification</td>
</tr>
<tr>
<td><strong>6</strong></td>
<td>Smart Therma-glove System for Raynaud’s Disease</td>
</tr>
<tr>
<td><strong>7</strong></td>
<td>Using Interactive 3D Software to Create Manipulatable Human Figures for Body Perception Research.</td>
</tr>
<tr>
<td><strong>8</strong></td>
<td>Using dynamical models to test experimental hypotheses in repeated-measure designs</td>
</tr>
<tr>
<td><strong>9</strong></td>
<td>Minding the Details: Use of Mind-Mapping Software to Teach Psychology and Probability</td>
</tr>
<tr>
<td><strong>10</strong></td>
<td>Narrowing in on the Timing of Semantic Representation in the Brain using Multiband fMRI during Reading</td>
</tr>
<tr>
<td><strong>11</strong></td>
<td>Continuous Response Interface Boosts Online Donations</td>
</tr>
<tr>
<td><strong>12</strong></td>
<td>The Political Semantic Space from 2000 to Present</td>
</tr>
<tr>
<td><strong>13</strong></td>
<td>Ti-Med</td>
</tr>
<tr>
<td><strong>14</strong></td>
<td>Peltier Effect Driven Fluidic Cooling Vest for Multiple Sclerosis</td>
</tr>
<tr>
<td><strong>15</strong></td>
<td>Bluetooth Insulin Pump System to Reduce Social Stresses of Insulin Pump Use</td>
</tr>
<tr>
<td><strong>16</strong></td>
<td>Careful what you post: Pronoun choice exacerbates emotions in tweets about terrorism</td>
</tr>
<tr>
<td><strong>17</strong></td>
<td>Improved EDA Measurement Device</td>
</tr>
<tr>
<td><strong>18</strong></td>
<td>User’s Experience and Strategy Use with Pokémon Go</td>
</tr>
<tr>
<td>19</td>
<td>Struggling Adult Readers Spend More Time Choosing to Learn, but Achieve Comprehension Strategy Learning Comparable to Normal Readers</td>
</tr>
<tr>
<td>20</td>
<td>ECG Drowsiness Detection System</td>
</tr>
<tr>
<td>21</td>
<td>Force Sensing Insoles with Therapeutic Replacement Padding</td>
</tr>
<tr>
<td>22</td>
<td>Web-based Experiment Management</td>
</tr>
<tr>
<td>23</td>
<td>Complexity Matching and Linguistic Convergence Across Bilingual Interlocutors</td>
</tr>
<tr>
<td>24</td>
<td>Towards Understanding Lexical Psychological Phenomenon and Mental Engagement via the Stochastic Model of Buddhism Psychology</td>
</tr>
<tr>
<td>25</td>
<td>Evaluation of Dialogues and Trialogues in ITS</td>
</tr>
<tr>
<td>26</td>
<td>AXIS: Generating Explanations at Scale with Learner sourcing and Machine Learning</td>
</tr>
<tr>
<td>27</td>
<td>Learning from errors by doing in online math tutoring system</td>
</tr>
<tr>
<td>28</td>
<td>A Testing Utility for AutoTutor</td>
</tr>
<tr>
<td>29</td>
<td>Topic Network Analysis on a Summarization Corpus</td>
</tr>
<tr>
<td>30</td>
<td>Data collection and sharing in AutoTutor System using Experience API (xAPI)</td>
</tr>
<tr>
<td>31</td>
<td>Conversation-based Assessment Systems created with AutoTutor Script Authoring Tool for Assessment</td>
</tr>
<tr>
<td>32</td>
<td>Learning from OperationARA: A Game-like version of AutoTutor</td>
</tr>
</tbody>
</table>
 KEYNOTE

Dr. Arthur Graesser
Professor, Department of Psychology
Co-Director, Institute for Intelligent Systems
The University of Memphis

Twenty Years of AutoTutor: Computer Agents Help Students Learn by Holding Conversations in Natural Language

Dr. Graesser is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis, as well as an Honorary Research Fellow at University of Oxford. He received his Ph.D. in psychology from the University of California at San Diego. His research interests question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, problem solving, memory, emotions, artificial intelligence, computational linguistics, and human-computer interaction. He served as editor of the journal Discourse Processes (1996–2005) and Journal of Educational Psychology (2009–2014), as well as presidents of 4 societies, including Society for Text and Discourse (2007-2010), the International Society for Artificial Intelligence in Education (2007-2009), and the Federation of Associations in the Behavioral and Brain Sciences (2012-13). He and his colleagues have developed and tested software in learning, language, and discourse technologies, including those that hold a conversation in natural language and interact with multimedia (such as AutoTutor) and those that analyze text on multiple levels of language and discourse (Coh-Metrix and Question Understanding Aid -- QUAID). He served on OECD expert panels on problem solving, namely PIAAC 2011 PSTRE, PISA 2012 Complex Problem Solving, and PISA 2015 Collaborative Problem Solving (chair).
PRESIDENTIAL SYMPOSIUM

Dr. Brendan T. Johns
Communicative Disorders and Sciences Department
University at Buffalo

Content-Driven Machine Learning: Using Lexical Variability to Optimize Models of Natural Language

Brendan T. Johns is an Assistant Professor at the University at Buffalo in the Communicative Disorders and Sciences Department. He obtained his Ph.D. in 2012 from Indiana University. His research is based on knowledge acquisition and his work exploits large scale computational models of language and memory. He has published on a range of topics including knowledge acquisition, natural language processing, human memory, perceptual-lexical integration, language-based deficits in Alzheimer’s disease, and computational statistics. He has won several awards including the Best Article Award from the Canadian Society for Brain, Behavior, and Cognitive Science, the Marr prize from the Cognitive Science Society, and the Castellan award from the Society for Computers in Psychology.

Dr. Evimaria Terzi
Computer Science Department
Boston University

Computational methods for team formation

Evimaria Terzi is an Associate Professor at the Computer Science Department at Boston University. Before joining BU in 2009, she was a research scientist at IBM Almaden Research Center. Evimaria has received her Ph.D. from University of Helsinki, Finland and her MSc from Purdue University. Evimaria is a recipient of the Microsoft Faculty Fellowship (2010) and the NSF CAREER award (2012). Her research interests span a wide range of data-mining topics including algorithmic problems arising in online social networks, social media and recommender systems.
Using R to explore ERP data

Chi-Lin Yu, National Cheng Kung University
David Causeur, Agrocampus Ouest
Ching-Fan Sheu, National Cheng Kung University

Event-related potentials (ERPs) are recordings of electrical activity along the scalp time-locked to stimulus or response events. ERP data provide researchers with important information about the chronology of mental processes. In R, an open-source, software environment for statistical computing and graphics, several packages (e.g., ERP, erpR, erp.easy, eegkit) are available for users to explore and analyze ERP data. Unfortunately, functions implemented in these packages suffer from being too simplistic, on the one hand, and too complex for users, on another, to conduct a comprehensive, graphical ERP data exploration. The purpose of the present work is to provide a single R function implementing the Grammar of Graphics plots popularized by Wickham (2009). The function allows users to control the process of data visualization through changes of one or several plotting options without having to worry (too much) about data manipulation. For example, users can decide to plot on the ERP 10/10 system simply by assigning an appropriate value to one of many options in the R function. For identifying (graphically) significant time points where groups of ERP curves might differ, the current function interfaces with the factor-analytic multiple comparisons procedures implemented in the “ERP” package (Causeur, Chu, Hsieh, & Sheu, 2012; Sheu, Perthame, Lee, & Causeur, 2016). In addition to demonstrating the graphical capabilities of the R function by exploring several built-in ERP datasets in R packages, we also note that the current function can be extended to handle data arising from instruments such as Near-infrared spectroscopy (NIRS).

An R package to conduct LME based item-level analysis of performance in artificial grammar experiments

Matthew T. Cook, University of Manitoba
Chrissy M. Chubala, University of Manitoba
Randall K. Jamieson, University of Manitoba

To simplify the problem of studying how people learn natural language, researchers use the Artificial Grammar Learning task. In this task, participants study letter strings constructed according to artificial grammar rules and must subsequently discriminate grammatical from ungrammatical test strings. Whereas data from these experiments are usually analyzed by comparing mean discrimination performance between experimental conditions, the practice discards information about individual items and subjects that could otherwise help to uncover the particular linguistic features guiding participants’ judgements. However, individual item analyses are tedious to compute, error-prone, and ill-defined in the literature. Moreover, the data violate the assumption of independence underlying standard linear regression models. To solve these problems, we present a suite of free programs for researchers studying artificial grammar learning. The programs’ intuitive web-based user interface
allows researchers to generate strings from a database of published grammars, compute item-level statistics (e.g., Levenshtein distance) for each string, and conduct Linear Mixed Effects (LME) analyses on their data. The LME analysis is an improvement over Lorch and Myers’ (1990) method of controlling Type 1 error inflation that is commonly used for repeated measures regression analyses (e.g., Kinder & Assmann, 2000). Finally, the software can generate a number of graphical representations to facilitate analysis and interpretation. The program suite is available as both a web-based Shiny app and a command-based R package. We hope the ease and availability of these tools will encourage researchers to take full advantage of their datasets in the study of Artificial Grammar Learning.

**Silky - A rich, graphical alternative to SPSS, which unifies graphical statistics software and the R command line**

Jonathon Love, University of Tasmania
Damian Dropmann, Centre for Liberated Scientific Software

The field of statistical analysis is divided between graphical software, such as SPSS, and command line programming languages such as R. Graphical software has the advantage that it is easy to use and accessible to a broad audience, however developing rich user interfaces and interactive analyses can be difficult for people without a background in software development. In contrast, R, although easy to develop sophisticated analyses, its command line interface can be a challenging environment for beginners. This talk introduces the Silky project, a community driven statistics project which bridges the gap between the graphical and command line communities. The project provides SilkyUI, a rich interactive statistical environment which is a suitable alternative to SPSS. Of particular note, SilkyUI is able to provide the underlying R code for each analysis, making it easy for users to make the jump to the command line. Additionally, the Silky project provides frameworks that make it easy to create and publish R packages which are useable from both an interactive graphical environment like SilkyUI, or the command line. In this way, the Silky project makes R more approachable and invites graphical statistics users to the command line, and at the same time makes it easy for R developers to provide rich interactive analyses for the graphical statistics community. This talk introduces Silky, demonstrates its use and feature set, and shows how existing analyses from R can be adapted to be rich, interactive, and driven from an accessible user-interface. Silky is available from silkyproject.org

**Using complicated n-gram smoothing algorithms efficiently in R**

David W. Vinson, University of California Merced

Massive natural datasets are now widely available for public use. As such, behavioral scientists have the opportunity to validate laboratory based theories ‘in the wild’. Many of these datasets consist of millions of cases of linguistic content such as online reviews and [micro]blog posts. Given the size of these datasets, even the simplest language models, such as n-gram analyses, require considerable computational power to use. The necessary computational requirements impose soft limits—available only to those trained in computational efficiency—to these rich datasets even though they are free to use. In an effort to bridge computational efficiency with behavioral research agendas, in previous work we developed the R package, cmscu, a replacement to the standard DocumentTermMatrix function in R’s tm package. cmscu, uses a probabilistic hash function (a count min Sketch) that affords efficient use of even the most sophisticated n-gram algorithms. Here, I will discuss how cmscu can be used to
implement some of the most sophisticated n-gram algorithms to date including modified Kneser-Ney smoothing. We hope to provide behavioral scientists access to tools and data previously out of reach.

8:00am – 9:00am  Fairfax B

Games and Web

Use of Warnings for Instructing Users How to Detect Phishing Webpages

Aiping Xiong, Purdue University
Robert W. Proctor, Purdue University
Ninghui Li, Purdue University
Weining Yang, Purdue University

The ineffectiveness of phishing warnings has been attributed to users' poor comprehension of the warning. However, the effectiveness of a phishing warning is typically evaluated at the time when users interact with a suspected phishing webpage, which we call the effect with phishing warning. Nevertheless, users' improved phishing detection when the warning is absent -- or the effect of the warning -- is the ultimate goal to prevent users from falling for phishing scams. We conducted an online study to evaluate the effect with and of several phishing warning variations, varying the point at which the warning was presented and whether procedural-knowledge instruction was included in the warning interface. The current Chrome phishing warning was also included as a control. 360 Amazon Mechanical-Turk workers made decisions about 10 login webpages (8 authentic, 2 fraudulent) with the aid of warning (first phase). After a short distracting task, the workers made the same decisions about 10 different login webpages (8 authentic, 2 fraudulent) without warning. In phase one, the compliance rates with two proposed warning interfaces (98% and 94%) were similar to those of the Chrome warning (98%), regardless of when the warning was presented. In phase two (without warning), performance was better for the condition in which warning with procedural-knowledge instruction was presented before the phishing webpage in phase one, suggesting a better of effect than for the other conditions. With the procedural knowledge of how to determine a webpage's legitimacy, users identified phishing webpages more accurately even without the warning being presented.

Human Trust in Automation in a Phishing Context

Jing Chen, New Mexico State University
Robert W. Proctor, Purdue University
Ninghui Li, Purdue University

Many previous studies have shown that trust in automation mediates the effectiveness of automation in maintaining performance, and one critical factor that affects trust is the reliability of the automated system. In the cyber domain, automated systems are pervasive, yet the involvement of human trust has not been studied extensively as in other domains such as transportation. In the current study, we used a phishing email identification task (with a phishing detection automated assistant system) as a testbed to study human trust in automation in the cyber domain. More specifically, we systematically investigated the influence of “description” (i.e., whether the user was informed about the actual
reliability of the automated system) and “experience” (i.e., whether the user was provided feedback on their choices), in addition to the reliability level of the automated phishing detection system. These factors were varied in different conditions of response bias (false alarm vs. misses) and task difficulty (easy vs. difficult), which were found may be critical in a pilot study. Measures of user performance and trust were compared across different conditions. The measures of interest were human trust in the warning (a subjective rating of how trustable the warning system is), human reliance on the automated system (an objective measure of whether the participants comply with the system’s warnings), and performance (the overall quality of the decisions made).

**Can playing video games produce flow experiences?**

J. Scott Jordan, Illinois State University  
Devin Gill, University of Utah

The present study investigated McGonigal’s (2011) claim that video games give rise to flow experiences; what Nakamura and Csikszentmihalyi (2002) define as one’s sense of deep engagement in an intrinsically rewarding activity. Over two blocks of 40 trials, participants played a video game in which they used key presses to make a stimulus move back and forth across the computer screen. After a few passes across the screen, the stimulus vanished and participants indicated (1) the perceived vanishing point, and (2) either their degree of sensed agency over the stimulus, or the amount of time they felt had passed during the trial. Participants completed a flow questionnaire after each block. While their trial-to-trial sense of agency correlated with both the magnitude of the perceived vanishing point and flow measures, their trial-to-trial sense of passed time, did not. Explanations are given for why these data only partially support McGonigal’s (2011) claims.

**A gamified mobile web app for real-time in-lecture quizzes**

Stian Reimers, City University of London

Last year I presented a new learning technology where students could send answers to multiple-choice quizzes run in lectures via text message. This year I introduce a mobile web application for running quizzes on any networked device. The basic idea draws on learning and memory research, with distributed retrieval practice at its core. With the app, students see quiz questions appear on their screen in real time as a lecturer presents them to the class. They select the correct answer for each question, and at the end of the quiz, they see their score. Gamified elements -- like the gaining of badges and trophies -- offer encouragement. Students can track their performance across quizzes from week to week on a leaderboard, and receive simple tailored feedback on their performance highlighting their relative strengths and areas for improvement. The new version has several advantages over the text-message version: It costs nothing for the user; It does not rely on having a mobile signal; It is much more straightforward and intuitive to use; It can provide on-screen tailored feedback to the student. I will introduce the technology, its rationale and implementation, and demonstrate how it can potentially be used to enhance learning and student engagement.
Computational Semantics

They Got Inside my Head: Using Knowledge Representations Derived from Text to Analyze Open-Ended Survey Responses

Anne Lippert, The University of Memphis

Open-ended response data from surveys is a rich, varied source of information and is often seen as offering a direct view into a respondent’s mind. Yet, compared to closed-ended response data, open-ended response data is alarmingly under collected and under analyzed. When open-ended survey data is collected, its analysis almost exclusively requires human coding, which can be tedious and time consuming. I present an alternative, semi-automated approach, Open-Ended Response Networks (ORNets), that draws upon cognitive psychology’s extensive work in knowledge representation to extract and represent respondent attitudes embedded within open-ended response data. I make two crucial contributions with this work: First, ORNets represent the first semi-automated network approach for analyzing open-ended response data that is grounded in a psychologically valid representation technique- the Pathfinder scaling method. Second, ORNets represent open-ended response data as semantic networks and thus allow researchers to analyze qualitative data using the quantitative approaches of network theory. This article details the construction of ORNets and highlights how they make analyzing open-ended responses easier and more informative. I illustrate these innovations by reviewing research that has recently and successfully used ORNets to analyze open-ended response data.

Automated computation of continuous semantic categories from unambiguous labelled exemplars

Chris Westbury, Department of Psychology, University of Alberta
Geoff Hollis, Department of Psychology, University of Alberta

Automated computation of semantic categories has largely relied upon extrapolating human judgments (e.g. ratings of imageability, gender, or affect), which are often of low reliability. We present an algorithm for bootstrapping semantic categories computationally that is largely independent of human decisions, since it relies only on unambiguous binary decisions about category membership. For example, while we might dispute how masculine or feminine we should rate the words 'giraffe', 'cloud', or 'president', there will be perfect agreement on whether 'queen', 'boy', 'actress', 'lady', and 'Thomas' should be rated as either masculine or feminine. Our method begins by identifying unambiguous exemplars of a category, which we categorize as '1' or '0'. We average together the skip-gram vectors for each pole of these exemplars, and find the words with vectors that are most similar to those average vectors in order to generate more unambiguous exemplars. This process may be repeated, averaging together an ever-growing set of unambiguous exemplars. We then use logistic regression across the vector values to build a regression model to predict category membership across this set, resulting in a model that ranks category membership in a continuous manner. This process of selection of unambiguous exemplars may be repeated using the values from the regression model in order to further 'tune' the model. We show that this simple method has high face validity and good
convergence with human ratings on a range of categories, and demonstrate how the method can be extended to create new semantic categories.

**Automated Summary Assessment: Comparison of LSA Similarity with Varied References**

Haiying Li, Rutgers University  
Zhiqiang Cai, The University of Memphis  
Art Graesser, The University of Memphis

Latent semantic analysis (LSA) similarity is a popular feature for automated summary assessment. LSA similarity compares the extent to which the target summary is similar to reference summary. The frequently-used referenced summaries included expert summaries (good or multi-scales), source texts, and crowdsourcing-based summaries generated by all summary writers. Previous research showed that LSA similarity with these references predicted human-graded summaries well when it was used alone or along with other linguistic and discourse features. However, it is unclear whether crowdsourcing-based LSA similarity could better predict human-graded summary scores than others. In addition, as LSA similarity with crowdsourcing-based summaries as reference summaries was a robust predictor for human summary score, does it predict human summary score better than crowdsourcing-based summaries with good, intermediary, or poor as references, respectively? If these are true, crowdsourcing-based summaries could replace expert summaries to reduce the time-consuming step of human summary grading and expert summaries. The current study compared the performance of LSA similarities with 8 types of references: source texts, expert summaries (good, intermediary, poor), and crowdsourcing-based summaries (all, good, intermediary, poor) within and across the source texts. Results showed that all of LSA similarities were significantly correlated, but only all crowdsourcing-based LSA similarity was highly correlated with other LSA similarities and human-graded summaries. This implies that crowdsourcing-based LSA similarity could be utilized in automated summary assessment as humans perceive the construct of summary grading. This approach will promote research on automated summary assessment and expand its application to small-scale summary grading (e.g., school teachers).

**Cross-situational learning in a Zipfian environment**

Andrew T. Hendrickson, University of Adelaide  
Amy Perfors, University of Adelaide

How do children learn words when one word can refer to many possible referents in any given scene? One popular theory is that they can leverage the statistics of word usage across many different scenes in order to isolate specific word meanings (Yu & Smith, 2007). Both adults and children have shown impressive learning in this kind of cross-situational learning paradigm, but relatively little is still known about how well it scales to real language. Some have suggested that when words follow a Zipfian distribution (as they do in natural language), a full lexicon should not be learnable because of the ambiguity that arises from many low-frequency words that are only observed a few times (Vogt, 2012). In this work, we show the opposite: when the distribution of words matches natural speech -- i.e., when the distribution of words is Zipfian and short words occur more frequently -- adults show improved learning in cross-situational contexts. Over a series of experiments, we show that this effect extends beyond the high-frequency words: when matched for word frequency, Zipfian distributions...
produce better learning than a uniform distribution. The Zipfian advantage occurs even when items are shown one-by-one, suggesting that it emerges out of more general properties of human cognition rather than because the high-frequency words are especially useful for disambiguation.

The Associative Boost in Semantic Priming: May It Rest in Peace

Curt Burgess, University of California, Riverside
Sarah Maples, University of California, Riverside

The notion that words that are both semantically and associatively (SA) related show more priming than words that are just semantically (S) related has been attributed to a process known as “associative boost” (Moss, Hare, Day & Tyler, 1994, and others) and more recently in a statistical model using a large text corpus (Ghanem, 2015). Early research with the HAL model demonstrated that the semantic distance metric could serve as an objective measure of similarity. When S pairs and SA pairs showed the same relatedness effect (priming in the model), there was no difference in priming with human subjects (Lund, Burgess & Atchley, 1995). In other words, no associative boost. The associativity of word relationships has been thought to be represented by lexical co-occurrence frequency, however, this effect obtains only when the word pairs are contextual neighbors in the first place (Lund, Burgess, & Audet, 1996). In the current experiments, we investigate the relationship between neighborhood size and conditional probability (also thought to be a proxy for word associativity) and the effect of the size of the neighborhood on this relationship. We test this with the original USENET corpus and several much larger corpora. We conclude that the utility of the “associative boost” was more a function of getting neural network models to work in the absence of more informative statistical information that is part of the language stream and is encoded in these contextual learning models.

The Center for Data on the Mind: Helping cognitive scientists tackle the promise and challenges of big data to explore behavior and cognition

Alexandra Paxton, University of California, Berkeley
Thomas L. Griffiths, University of California, Berkeley

Big data offers cognitive scientists unparalleled opportunities to explore human behavior and cognition, but it poses its own set of unique challenges. The Center for Data on the Mind is a new, community-focused initiative aimed at equipping cognitive scientists with the mindset and resources to take full advantage of these opportunities. From highlighting unique datasets and tools to profiling inspirational projects, our goal is to encourage new, data-rich, naturalistic studies of human behavior and cognition with big data.

9:15am – 10:45am Fairfax B

Eye tracking and Classification

Gaze Path Segmentation with the Beta Process Autoregressive Hidden Markov Model
The first stage of analyzing eye-tracking data is commonly to code the data into sequences of fixations and saccades. This process is usually automated using simple, predetermined rules for classifying ranges of the time series as either of the two types, such as “the rate of change of the fixation location is less than the threshold, then code as a fixation; otherwise code as a saccade.” More recent approaches have attempted to include additional eye-movement categories in automated parsing algorithms, particularly glissades, by using time-varying, data-driven thresholds. We will describe an alternative approach using the beta-process autoregressive hidden Markov model (BP-AR-HMM). The BP-AR-HMM offers two main advantages over existing frameworks. First, it provides a statistical model for the classification rather than just a single estimate. Second, the BP-AR-HMM uses a latent process to model the number and nature of the types of eye-movements and hence is not constrained to the two (or three) predetermined types. We applied the BP-AR-HMM to a subset of data from the DIEM project, a freely available large scale corpus of thousands of scan paths recorded from participants watching short video clips. Based on these data, the BP-HMM was able to identify saccades, fixations, and third category that approximately corresponds with glissades (as well as blinks/errors/etc.).

**A minimal simulation model of foveated eye movements and visual search**

Shane T. Mueller, Michigan Technological University
Yin Yin Tan, Michigan Technological University
Hannah North, Michigan Technological University
Kelly Steelman, Michigan Technological University

We describe the Eye Movement Minimal Model-Modified (EM4), a lightweight minimally-sufficient model of eye movements that accounts for visual search times in several distinct paradigms. The model allows visual search to be guided by probe-item similarity in different foveal zones, which enables the model to be used as a front-end for various models of visual saliency. We apply the model to visual search and change detection paradigms, to examine its utility and accuracy.

**ROC analysis of eye movements on complex images mediated by linguistic context**

Stephanie Huette, The University of Memphis
Jeffrey Viaud, The University of Memphis

Models of visual saliency are used to predict eye movements and where attentional resources are likely to be allocated (Itti, Koch, & Niebur, 1998). However, eye movements change as a function of task (Yarbus, 1967) which entails use of linguistic and conceptual knowledge. To understand the fundamentals of how a pragmatic context mediates eye movements, a study was designed using negation as the primary manipulation. We predicted negation functions as a learned pragmatic context to drive eye-movements toward alternatives in a natural scene. Participants passively listened to a vignette while viewing a grey screen, and then the corresponding image was shown for 15 seconds. An ROC eye movement analysis shows how many fixations on average hit places that were predicted to be salient from a GBVS model of saliency (Harel, Koch, & Perona, 2007). The results demonstrate salience of an image differ between language conditions and images but in this analysis, contrary to other
analyses, ROC shows little between-subject variability.

**Using Best/Worst Scaling to Collect Semantic Judgments**

Geoff Hollis, University of Alberta

Both psycholinguistics and natural language processing have recently shown an interest in compiling high-coverage datasets for semantic properties of words using crowdsourcing tools (e.g., Kiritchenko & Mohammad, 2016; Brysbaert, Warriner, & Kuperman, 2014; Warriner, Kuperman, & Brysbaert, 2013; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). However, these types of experiments are costly to run. We introduce a rating format called best/worst scaling and describe how it can be used to efficiently collect human judgments for tens of thousands of items. We present the results of simulation studies that 1) establish the validity of this rating format for inferring latent semantic values of words and 2) provide recommendations for experimental design that helps improve data quality.

**Vocal emotion classification based on self-organizing map algorithms**

Xiaowei Zhao, Emmanuel College  
Zhi-Yong Tao, Harbin Engineering University  
Ya-Xian Fan, Harbin Engineering University

How to correctly recognize the emotions expressed in someone’s vocal communication? This has been a challenging interdisciplinary problem actively studied in several fields, including but not limited to psychology, acoustics and machine learning. Recently, Liu and Pell (2012) developed a database of vocal emotional stimuli in Mandarin Chinese, which provides the acoustic features and human classification rating of a set of pseudo sentences with various emotions. In the present study, we analyzed the data from the Liu & Pell database based on the Self-Organizing Map (SOM), an unsupervised neural network model that can project the inherent structure of the emotional stimuli onto a 2-D space based on their acoustic features. In addition, a set of variants of SOM (e.g. the supervised SOM and the Learning Vector Quantization algorithms) were used to obtain optimal machine classification results that were comparable to human recognition rates. Based on them, the prototypes (typical acoustic features) of each emotional category were also extracted.

**The QC-methodology: adapting the Q-methodology for analyzing Concepts through Categorical data**

Margot Brard, Applied Mathematics Department, Rennes, France  
Sebastien Le, Applied Mathematics Department, Rennes, France

In this presentation we will show how the Q-methodology can be twisted to understand a complex concept through a large set of stimuli. In the QC-methodology, each subject has to sort a given set of stimuli in predefined groups. These groups are either defined as the concept of interest and its negation, or as a gradient of the concept of interest. This kind of experiment generates intrinsically categorical data that have to be analyzed as such. We will show how to analyze this kind of data in order to get a graphical representation of the stimuli such as the distance used to represent the stimuli expresses the way they have been perceived in terms of concept. In order to corroborate the methodology, it was tested to understand the concept of innovation. The subjects were asked to assign
each 39 stimuli to the category that most reflected their viewpoint: innovative or not innovative. This experiment revealed several visions of innovation, which have been linked to textual descriptions. In the presentation, we will show how data have been collected with the Holos environment and analyzed with the SensoMineR package.

9:15am – 10:45am

SYMPOSIUM

Twenty years of Internet-based research at SCiP: Surviving concepts, new methodologies

Organizer: Ulf-Dietrich Reips
Discussant: Christopher Wolfe

It was 20 years ago, at SCiP in Chicago in 1996, that for the first time methodologies for web-based research were introduced, namely on Internet-based experimenting. With the present symposium proposal containing talks by authors from four countries we would like to review the methodologies and technologies in Internet-based research that have evolved since and will present some applications on the edge of current scientific frontiers. In the first talk, John Krantz and Ulf-Dietrich Reips will rigorously review current research examples in light of methodologies and principles of Internet-based research as they were developed during the first years, when researchers using the new method had to carefully meet the then many questions raised by peers. They diagnose a certain laxness in current Internet-based research practices and propose best practices. The second talk, by Stefan Stieger and Ulf-Dietrich Reips from the University of Konstanz in Germany and David Lewetz from the University of Vienna in Austria investigates how recent mobile technology, i.e. smartphones and smartphone apps, can be used to measure pace of life in its relation to stress. The authors report on an experience sampling approach with two types of measurement: self-report and a tapping task on the smartphone. Smartphone tilt as a possible posture-related indirect measure of well-being is the topic of the third talk by Tim Kuhlmann, Stefan Stieger and Ulf-Dietrich Reips. They report on a smartphone app study that assessed well-being and behavioral variables twice a day across a span of three weeks and also explore correlations with personality. The authors discuss challenges in analyzing data from smartphone sensors and issues in taking sensor data as a proxy of other behaviors. Data collected in Internet-based research about the process by which the data were collected are called paradata. Birk Diedenhofen and Jochen Musch present PageFocus, a JavaScript that allows to determine if and how frequently participants abandon test pages by switching to another window or browser tab. They conducted two experimental validation studies to validate whether PageFocus can detect and prevent cheating in unproctored online tests. The final talk, by Ulf-Dietrich Reips, Stefan Stieger, Sven Heinrichs and Alex de Spindler, picks up on the topic of smartphone based studies and presents a generic smartphone app configurator for mobile assessment studies. Longitudinal data logging and the combination of smartphone sensor data with self-report data over locations and time provide unprecedented options for fine-grained longitudinal research. The combination of powerful mobile computerization in small devices permanently connected with broad-bandwidth Internet and advanced Internet-based research methodology are a powerful yet logical development that began 20 years ago. Former SCiP president Christopher Wolfe will serve as a discussant for the symposium.
Update on the health and use of the web for psychological research

John H. Krantz, Hanover College
Ulf-Dietrich Reips, University of Konstanz

The first papers that introduce the conducting of psychological research on the web were presented at this conference 20 years ago (e.g., Krantz, Ballard, & Scher, 1996; Reips, 1996). Since that time, there has been an explosive increase in the number of studies that use the web for data collection as can be seen on sites such as Krantz (1996), Reips (1995, 2001), Reips & Lengler (2005) that post links to studies. As such, it seems a good time, 20 years on, to examine the health and best practices of research on the web. While the rapid increase in the use of the web indicates that the findings collected off of the web are deemed valid, a major concern of the early researchers, it might also suggest that the research method is getting less examination that it ought to. While many studies are well conducted, it is the experience of the authors, that along with the common usage of the web for research, there is also a certain laxness in the use of the web leading to ill-conceived studies being posted on the web. Several examples, denuded of identifying information, will be presented, and best practices for web research will be reviewed including reference to emerging online methodologies. Examples will range from simple where the title does not communicate well to a general audience to much more severe where the basic design, structure and concept of the study are problematic.

The measurement of pace of life: Results from an experience sampling smartphone app study

Stefan Stieger, University of Konstanz
David Lewetz, University of Vienna
Ulf-Dietrich Reips, University of Konstanz

People’s pace of life is moving into the focus of scientific research in many different research fields (e.g., psychology, ecology, medicine). In psychology, an overly high pace of life is often associated with health issues such as burnout, reduced life satisfaction, increased psychological stress, and others. Interestingly, pace of life is often measured indirectly (e.g., walking speed, punctuality of clocks) and data is analyzed on an aggregated level (e.g., culture). Studies measuring pace of life on an individual level are rare. Therefore, we conducted a smartphone app study to assess pace of life with two different measures by using an experience sampling approach (i.e., two measurements per day for three weeks; 6,000+ single assessments). Measures were 1) self-report (“How fast do you currently experience your life?, visual analogue scale from very slow to very quick) and 2) a tapping task, which asked participants to tap on the smartphone screen in the rhythm of one’s current pace of life. Furthermore, we assessed several potential correlates with pace of life such as well-being, psychological stress, and sleep duration as well as two non-reactive indicators of pace of life (time to fill in the smartphone app questionnaire, time to react to the reminder). Results: with self-reported pace of life it is possible to successfully replicate past research (e.g., the higher the psychological stress the higher the experienced pace of life). Furthermore, the non-reactive tapping task incrementally explains further variance to show a more complete picture of what it psychologically means to have variations in one’s pace of life.

Smartphone tilt as a measure of well-being? Results from a longitudinal smartphone app study
Tim Kuhlmann, University of Konstanz
Ulf-Dietrich Reips, University of Konstanz
Stefan Stieger, University of Konstanz

Smartphones offer new types of measurement, such as the indirect measure of posture via the device’s tilt. The current study investigated subjective well-being in association with personality and sleep quality using a smartphone app based experience sampling design. A total of 98 participants (age M = 22.9 years, SD = 6.7) downloaded a smartphone app that assessed well-being and other behavioral variables twice a day across a span of three weeks (level 1). Additionally, participants filled in a questionnaire assessing personality variables such as Narcissism and Extraversion at the end of the study (level 2). Multilevel analyses revealed a negative association between well-being and the objectively measured tilt of the smartphone while answering the items (d = .14, p < .001). Other effects, e.g. lower well-being scores in the morning as compared to the evening and negative effects of critical life events (d = .35, p < .001) and nightmares (d = .39, p < .001) on the subjective well-being, were found. The current study validates previous research on well-being in a longitudinal design. Sensor data did show a theoretically predicted association with well-being in the current study, but the effect appears more complex given conflicting evidence from previous studies. The challenges in analyzing potentially beneficial objective data from smartphone sensors are discussed. This includes different ways to analyze the data as well as problems of implementing sensor data as a proxy of other behaviors.

**PageFocus: Using paradata to detect window switching behavior in online tests**

Birk Diedenhofen, University of Düsseldorf
Jochen Musch, University of Düsseldorf

The validity of unproctored online tests is threatened by participants who switch back and forth between the test page and other web pages or applications. To address this problem, we developed PageFocus, a JavaScript that allows to determine if and how frequently participants abandon test pages by switching to another window or browser tab. The script can easily be implemented and is compatible with all major browsers. To validate whether PageFocus can detect and even prevent cheating in unproctored online tests, we conducted two experimental validation studies. In a parallel lab and web study, we found that test-takers who were invited to cheat abandoned the pages of a knowledge test more frequently and achieved higher scores. Presenting a popup warning, however, successfully reduced cheating rates. With operating system data and the test-takers' self-reports as external criteria, a very high sensitivity and specificity was found for cheating detection based on the PageFocus script. Concurrent evidence from the lab and the web sample suggests that our lab results generalize to testing in online contexts. In a second study, the PageFocus script revealed that participants cheated more when performance-related incentives rather than performance-unrelated incentives were being offered. As expected, cheating was only detected for items that allowed to look up the solution using a search engine. The results of our validation studies suggest that PageFocus can help to improve data quality in unproctored online tests.
Is there anybody out there? Sleepsex.org, a longstanding hub for research and exchange on sexsomnia

Mike Mangan, University of New Hampshire

The Website Sleepsx.org went live in 1999. The purpose of the site was to gather and provide information on a then-obscure sleep condition called “sexomnia,” which involves sexual behavior occurring in sleep. At that time, there were only a handful of publications on this topic, the first appearing in print in 1996. These articles were housed in library stacks and difficult-to-access databases and concerned an unknown, undiagnosed, potentially devastating condition, occurring in a difficult-to-reach population. Over the past 17 years, the site has served as data gathering resource. Thousands of anecdotes have been collected and have formed the basis of published research. It has been a resource for persons who have experienced sexsomnia, either as unconscious initiators or recipients, for legal professionals dealing with cases on sexual misconduct ostensibly occurring in sleep, for media types looking for provocative stories, and for many others. Historically, the Website emerged in a context of burgeoning Web use. At that time, the development of Web-based research approaches and technologies were creating new possibilities for knowledge seekers and changing, fundamentally, how investigators might approach knowledge creation and dissemination. Sleepsex.org supported the research process, from initial idea generation and literature review, to data collection, to publication and the sharing of knowledge. The site continues to attract visitors from around the world with shared interests, questions, and concerns pertaining to sexual behavior in sleep.

Svendroid: A generic smartphone app configurator for mobile assessment studies

Ulf-Dietrich Reips, University of Konstanz
Stefan Stieger, University of Konstanz
Sven Heinrichs, Zurich University of Applied Sciences
Alexandre de Spindler, Zurich University of Applied Sciences

Smartphones offer new types of measurement via sensors such as GPS for location, accelerometer (orientation), gyroscope (rotation and twist), proximity (e.g. of ear), magnetometer, various light sensors, barometer, thermometer, pedometer, heart rate sensor, hall sensor, and even radiation sensor in some models. Continuous data logging and combination of sensor data with self-report data over locations and time provide unprecedented options for fine-grained longitudinal research. Consequently, there is a small but growing literature on smartphone-based studies. These studies have mostly used hard coded smartphone apps that were specifically programmed for the study at hand, usually they cannot be reused for other studies. On the other hand, universal app creators such as the MIT app creator are complex, require a steep learning curve and do not take into account researchers’ particular needs. Commercial attempts have been limited and are pricy. To date, there is no open source generic smartphone app configurator for mobile assessment studies that uses smartphone sensors as triggers and so we developed one that we present here. Svendroid allows researchers to configure and run its generic smartphone app for research purposes without programming knowledge. It offers a simple click-, type- and drag-and-drop interface to select from an array of triggers (random, location, time, social media event, event from another app) and measurements (GPS, background noise, self-report in various formats). Participants can be invited via Svendroid’s user management system or join publicly advertised studies after download of the participant app. Collected data can be
viewed and downloaded from a statistics section. We present the software and an example study. Svendroid will be made available as open source software available from svendroid.com and github.

11:00 am – 12:15 pm
Fairfax A

**Experimental Methods**

**lab.js -- A Modular Foundation for Browser-Based Experiments**

Felix Henninger, University of Koblenz-Landau & Max Planck Institute for Research on Collective Goods

With the growing capabilities of web browsers, an ever-increasing number of researchers rely on browsers to conduct experiments and surveys -- regardless of whether data collection occurs via the internet or within the laboratory. Considerable improvements in tooling notwithstanding, browser-based experiments still lag behind dedicated laboratory-based experimental software with regard to flexibility and ease-of-use. We introduce lab.js, an open source library designed to vastly simplify the realization of and data collection with browser-based experiments. It controls the succession of stimuli and input handling, leaving researchers to supply the structure and content of their studies. Through lab.js, we aim to make browser-based experiments more widely accessible, and to enable researchers with a basic knowledge of HTML and CSS and very limited experience with JavaScript to build browser-based web experiments. We are confident that researchers familiar with web development will be able to realize their studies more quickly and easily using our library. More experienced developers will find it extensible, adaptable, and thereby suitable even for complex experiments that previously could have only been built manually.

**Metronome LKM: An open source virtual keyboard driver to measure experiment software latencies**

Pablo Garaizar, Universidad de Deusto
Miguel A. Vadillo, University College London

Experiment software is often used to measure reaction times gathered using keyboards or other input devices. In previous studies, the accuracy and precision of the time stamps was assessed through several means: a) generating accurate square signals from an external device connected to the parallel port of the computer running the experiment software, b) triggering the typematic repeat feature of some keyboards to get an evenly separated of key press events, or c) using a solenoid controlled by a microcontroller to press the input device that will be used in the experimental setup (keyboard, mouse button, touch screen). Despite the advantages of some of these approaches in some contexts, they do not address the latencies caused by the experiment software. Metronome LKM provides a virtual keyboard to assess experiment software. Using this open source driver, researchers can generate key press events using high-resolution timers and compare the time stamps collected by the experiment software with those gathered by Metronome LKM (with a nanosecond resolution). Our software is highly configurable (keys pressed, intervals, SysReq activation) and runs on 2.6-4.4 Linux kernels.

**Capturing the Writing Process: Keystroke Logging in a Writing Tutor**
Automated Writing Evaluation (AWE) systems provide an environment for students to receive instruction and engage in deliberate writing practice with summative and formative feedback. Despite the accuracy of their scoring algorithms, however, the pedagogical elements of these systems have much room for improvement. System developers rely on natural language processing (NLP) methods to examine the quality of students’ written products while largely ignoring information about students’ behavioral processes. The current study used system logs of keystrokes to develop online assessments of students’ writing performance. The behavioral processes enacted by writers are important elements of writing skill; therefore, our aim was to determine whether we could assess and model the quality and linguistic properties of essays by calculating indices related to students’ typing behaviors. A corpus of prompt-based essays and keystroke log data were collected from 126 undergraduates. Both basic and time-sensitive keystroke indices were calculated to analyze the behavioral patterns enacted by the students. These indices provided information about writing processes at the aggregate level (e.g., number of backspaces) as well as information about how these behaviors unfolded over time (e.g., variation in pauses across time windows). Holistic scores and linguistic properties of essays were calculated using NLP tools. Results indicated that keystroke indices accounted for 76% of the variance in essay quality and up to 38% of the variance in the linguistic characteristics. These findings suggest that keystroke analyses can recover valuable information about writing, which may ultimately help to improve student models and feedback in computer-based learning environments.

Mousetrap: Free, open-source, and cross-platform mouse-tracking and analysis plug-ins

Pascal J. Kieslich, University of Mannheim
Felix Henninger, University of Koblenz-Landau

Mouse-tracking -- the analysis of mouse movements in computerized experiments -- is becoming increasingly popular in the cognitive sciences. Specifically, mouse movements are taken as an indicator of commitment to or conflict between choice options during the decision process. Using mouse-tracking, researchers are gaining a better understanding of the temporal development of cognitive processes in a growing number of psychological domains. In the current contribution, we present plug-ins that offer easy and convenient means of recording and analyzing mouse-movements in laboratory experiments. First, we introduce plug-ins and corresponding Python packages that add mouse-tracking to OpenSesame, a popular general-purpose graphical experimental builder for the social sciences. In contrast to existing mouse-tracking implementations, mousetrap seamlessly integrates with existing experimental software and, besides, does not require programming skills as it can be easily handled via its graphical interface. Thus, researchers can benefit from the core features of a validated experimental software package and the many extensions available for it (e.g., the integration with auxiliary hardware such as eye-tracking, and the creation of interactive experiments). Second, we introduce the mousetrap library for the statistical programming language R. This library can import, preprocess, aggregate, and visualize mouse trajectories and calculates a variety of established measures for
curvature, complexity, velocity, and acceleration of movements. Importantly, the data need not be collected using the plug-ins presented in the current article, but can, in principle, stem from any mouse-tracking implementation. All plug-ins as well as the associated software are cross-platform, open-source and available free of charge from https://github.com/pascalkieslich/mousetrap-os

11:00 am – 12:15 pm Fairfax B

New Measures: Social and Cognitive

Bayes factors for the Linear Ballistic Accumulator Model of Decision-Making

Nathan J. Evans, University of Newcastle
Scott D. Brown, University of Newcastle

Evidence accumulation models of decision making have led to advances in several different areas of psychology. These models provide a way to integrate response time and accuracy data and to describe performance in terms of latent cognitive processes. Testing important psychological hypotheses using cognitive models requires a method to select between different versions of the models which assume different parameters to cause observed effects. The task of model selection using noisy data is difficult, and has proven especially problematic with current model selection methods which are all based on parameter estimation. We provide a method for computing Bayes factors through Monte-Carlo integration for the linear ballistic accumulator (LBA; Brown & Heathcote, 2008), a widely used evidence accumulation model. Bayes factors are used frequently for model selection with simpler statistical models, and they allow model selection without parameter estimation. In order to overcome the computational burden of estimating Bayes factors via brute force integration, we exploit general purpose graphical processing units; we provide free code for this. This approach allows estimation of Bayes factors via Monte-Carlo integration within a practical time frame. Simulation studies show that this method of model selection performs well in realistic designs, outperforming the popular deviance information criterion.

Group communication analysis: A computational-linguistic framework for exploring conversational roles in online multi-party communication

Nia Dowell, The University of Memphis
Tristan Nixon, The University of Memphis
Zhiqiang Cai, The University of Memphis
Dr. Arthur Graesser, The University of Memphis

Communication, collaboration and the social co-construction of knowledge are critical 21st century skills in learning, working, and navigating our society. During collaborative activities, group members take on different roles within the discussion. Roles have distinct patterns in behavioral engagement style (i.e., active or passive, leading or following), contribution characteristics (i.e., providing new information or echoing given material), and social orientation (i.e., individual or group). Different combinations of these roles can produce characteristically different group outcomes, being either less
or more productive towards collective goals. In online collaborative learning environments, this can lead to better or worse learning outcomes for the individual participants. In this study, we present a novel approach for detecting emergent learner roles from the participants’ contributions and patterns of interaction. Specifically, we developed a group communication analysis (GCA) by combining automated computational linguistic techniques with analyses of the sequential interactions of online group communication. The GCA was applied to collaborative learning datasets. Cluster analyses and hierarchical linear mixed-effects modeling were used to assess the validity of the GCA approach and the influence of learner roles on student and group performance. The results indicate that learners’ patterns in linguistic coordination and cohesion are representative of the roles that individuals play in collaborative discussions. More broadly, GCA provides a framework for researchers to explore the micro intra- and inter-personal patterns associated with the participants’ roles and the sociocognitive processes related to successful collaboration.

**Context dependent measures of Synchrony in Team Sports**

Maurici A. López-Felip, University of Connecticut
Tehran J. Davis, University of Connecticut
Till D. Frank, University of Connecticut

Current models of Dynamical Systems Theory measure self-organization of team sports in terms of synchronization. Components of team systems (i.e., players) are considered as periodic phase oscillators whose degree of synchrony reflects the coordination of a team to achieve a common goal. Nonetheless, a main limitation of these models is that the collective behavior is context independent. In other words, N phase oscillators with phases theta-j for j=1,2,...N can be highly synchronized without this corresponding to a meaningful coordination dynamics relevant to the functional semantics of the game. Thus, the aim of this study was to develop a method of analysis sensitive to the contextual situations of the game and compare it to current models. Data were collected via GPS (10 Hz) for thirty minutes’ scrimmage. We analyzed the degree of synchronization for both teams in terms of the Kuramoto order parameter (cluster amplitude) \( r \), (high synchronization = 1) by measuring location data \((x,y\) pitch coordinates) over time. Team synchronization was measured in 10 plays of interest showing that changes on synchronization corresponded to events in the match. In addition, we analyzed whether a qualitative change in the team coordination predicts meaningful events in the match (e.g., change in ball possession, scoring opportunities, etc.).

By modeling Organism (athlete) and Environment (i.e., goals) as two coupled dynamical systems we show that context dependent models of social joint actions can have large implications for understanding intentional collective behavior. Implications for developing artificial social collectives will be discussed.
Examining the impact of tutor dialogue moves on knowledge in a Wizard of Oz system

Colin L. Widmer, Miami University
Christopher R. Wolfe, Miami University
Cecelia M. Favede, Miami University
Richard Smith, Miami University

Computer tutoring systems that interact with learners in natural language dialogues and guide learners to generate explanations of learned material produce reliable and powerful learning gains. These systems can take advantage of many dialogue acts, and identifying the specific impact different dialogue has on learning is an important research question for the development of automated tutoring systems. To investigate the influence of tutor dialogue moves on learning outcomes we developed a web-based Wizard of Oz system that allows a researcher to simulate an intelligent tutoring system with tightly controlled dialogue delivered to learners. In this study participants received a short lesson about breast cancer risk and screening from a computer avatar followed by an interactive dialogue with the avatar controlled by a researcher through the Wizard of Oz system. Participants received dialogues with one of four different patterns of dialogue moves: dialogue that only encouraged general elaboration about the lesson, dialogue that only requested specific information from the lesson, an alternating combination of both types of dialogue moves, or a control condition with no dialogue. All participants who received dialogue from the avatar showed increased knowledge, however the type of dialogue affected what knowledge increased. Participants who received prompts for specific information from the lesson showed better recollection of facts from the lesson, while participants who received dialogue encouraging elaboration showed a better understanding of the broader important gist of the lesson. In addition, we present detailed analyses of participant dialogues that reveal other subtle effects of dialogue differences.

11:00 am – 12:15 pm

**SYMPOSIUM**

**AutoTutor**

*Organizer: Xiangen Hu*
*Discussant: Danielle McNamara*

**Art History**

Danielle S. McNamara, Arizona State University

Art Graesser is well known for his cutting edge, and artistic, contributions across multiple areas of cognitive and learning sciences including knowledge representation, discourse processing, inference generation, conversation, question asking and answering, emotion, human computer interaction, serious games, and intelligent tutoring systems (to name a few). One major contribution to the field was his conception and development of AutoTutor, a pedagogical agent that could talk to and understand humans while they learned complex concepts such as physics. The creation of AutoTutor
was a major breakthrough in itself, but it also opened up doors that allowed the world to believe that such tutoring systems could be built, and they worked. This talk will discuss the historical significance of Art's work, focusing in particular on the development of natural language processing systems to undergird interactive, adaptive educational technologies.

Impact of Pedagogical Agents’ Conversational Formality on the Formality of Written Summaries

Haiying Li, Rutgers University
Art Graesser, The University of Memphis

English language learners received training on the rhetorical structures of compare-contrast and cause-effect informational genres and later wrote summaries of texts in these genres in addition to providing ratings and answering multiple-choice questions. We examined the impact of this training that manipulated conversational formality on the formality of the summaries that they wrote. The training consisted of an intelligent tutoring system with conversational dialogues in which a teacher agent, a student agent, and a human learner engaged in a conversation on these rhetorical structures. The learners were randomly assigned into one of the three conditions: (1) a formal condition in which both the teacher agent and the student agent spoke with a formal communication style, (2) an informal condition in which both agents spoke informally, and (3) a mixed condition in which the teacher agent spoke formally, whereas the student agent spoke informally. The results indicated that learners with a higher formality style at pretest wrote more formal summaries in the informal training condition than in the mixed condition. Thus, the informal conversational style helped them better understand the rhetorical structures and subsequently apply a formal style of writing. Follow-up analyses on five automated measures of language and discourse (via Coh-Metrix) also revealed that the cohesion of the summaries were influenced by some interactions between the learners’ prior formality style at pretest and the experimental manipulation of the agents’ formality. The article discusses the implications of these results for teachers and future research.

Understanding sentence and text level factors that affect performance on the Reading Strategy Assessment Tool

Joe Magliano, Northern Illinois University
Danny Feller
Laura Allen
Keith Millis
Danielle McNamara, Arizona State University

The Reading Strategy Assessment Tool (RSAT) was developed to assess important processes underlying reading comprehension. The assessment has students produce think aloud like responses as they comprehend texts, and computational algorithms score the responses on the extent that they generate bridging and elaborative inferences. Traditionally, bridging and elaboration scores are computed by aggregating over items. What is currently unknown is the effect that sentence- and text-level factors may have on the generation of these inferences. The goal of the present study is to explore sentence- and text-level factors that correlate with the bridging and elaboration scores at the item level. Preliminary analyses have been conducted comparing RSAT scores for bridging and elaboration with human judgments of those processes. These analyses reveal that the ability to detect these processes
is high and comparable to what has been reported at the person level. Analyses are ongoing and cannot yet be reported. Multilevel modeling will be used to assess the extent that sentence level (e.g., mean word frequency, causal connectivity) and text level (genre, cohecmextrix measures) factors are predictive of bridging and elaboration scores. The results of this study will inform versions of RSAT that can provide feedback at the item level.

AutoTutor Question Answering

Zhiqiang Cai, The University of Memphis
Xiangen Hu, The University of Memphis
Brent Morgan, The University of Memphis
Lijia Wang, The University of Memphis
Whitney Baer, The University of Memphis
Art Graesser, The University of Memphis

AutoTutor is a conversational intelligent tutoring system. The conversation is usually guided by a "hint-prompt-assertion dance, where AutoTutor asks hint and prompt questions to help students construct a good answer to a main question, with an assertion when the hints and prompts fail. In each conversation turn, AutoTutor gives feedback to a student input and selects a new question to ask. There has been considerable research on the importance of encouraging students to ask questions. However, constructing a question answering component in AutoTutor is a difficult task. Since AutoTutor is mostly an answer questioning system, picking out questions from answers needs an accurate question identification algorithm, which often causes a delay in the conversation due to the complexity of the algorithm. Because AutoTutor is a scripted conversation system, the answers to students' questions need to be pre-constructed. That means we need to know in advance what questions students might ask. Unfortunately, it is difficult for script authors (i.e., domain experts) to imagine students' questions. The most reliable way is to collect student questions during the conversation. However, our research indicates that students rarely ask questions, especially deep questions. In order to add a question answering component to AutoTutor that effectively helps students learn, we propose in this paper to add the following features to AutoTutor: (1) a "push-for-question" component that invites student questions; (2) a question modeling component that uses a peer student agent to model good question asking; (3) a question collection component that collects questions from real student conversations and presents the questions to AutoTutor script authors; (4) a question-answer authoring component that allows AutoTutor script authors to add questions and answers to AutoTutor systems.

Integrating the BRCA Gist intelligent discourse system into routine workflow in clinical settings

Christopher R. Wolfe, Miami University
Elyse E. Lower, University of Cincinnati

Previously we used AutoTutor LITE to develop an Intelligent Tutoring System (ITS) called BRCA Gist, the first ITS applied to lay people's medical decision making. It helps women understand genetic testing for breast cancer risk by engaging them in tutorial dialogues. BRCA Gist has modules on quantitative concepts, genetic testing, genetic risk, and the consequences of testing. An animated agent provides didactic information orally and through images, sounds, text, and brief animated sequences. The agent
poses questions and people respond by typing in a text box. The animated agent responds differently to each person depending on what they say with the goal of helping them form a good gist explanation in their own words. Previous randomized controlled experiments have demonstrated the efficacy of BRCA Gist. However, BRCA Gist had not previously been integrated into routine workflow in clinical settings. We present preliminary data from research at the University of Cincinnati Cancer Institute Comprehensive Breast Cancer Center on patient perceptions of BRCA Gist and discuss strategies for integrating a proposed next-generation system called BCCA (Breast Cancer Conversational Agent) into clinical practice. Women will interact with BCCA by (orally) talking to mobile devices such as tablets and smart phones. In addition to use as a stand-alone web-based resource for women, healthcare providers will hand patients tablets to interact with BCCA in the waiting room before or after the clinical encounter to help them understand complex concepts, make sound medical decisions, and make the most of precious time with doctors and genetic counselors.

<table>
<thead>
<tr>
<th>1:00 pm – 2:15 pm</th>
<th>Ballrooms in Level 3 of the Hynes Convention Centre</th>
</tr>
</thead>
</table>

**POSTER SESSION**

(1) *Given Names and Systemic Bias Replication through a Semantic Model*

Dawn Carroll, California State University, Fullerton

In the United States, there are approximately 195,709 inmates in federal custody (Federal Bureau of Prisons). The United States accounts for the citizenship of 77.6% of these inmates, 15.2% originate from Mexico, and the rest come from various parts of the world. 66.2% of inmates are non-Hispanic, while the rest identify as being of Hispanic origin. Over 71% of federal inmates are people of color (FBP). The general public’s belief of there being systemic racial bias in the United States prison system goes beyond the color of one’s skin. Some inmates report that having an “ethnic” or non-Eurocentric name can put them at a disadvantage during a court trial. In a previous study, my colleagues and I found that semantic tendencies were obtainable through gender-occupational bias with male and female given names (Bermudez, Carroll, Burgess, & Estep 2014). As a result, in this study, I have investigated all possible pairings of the semantic distances between masculine given names of various ethnic origins (Caucasian, Hispanic/Latino, and African-American) and a variety of courtroom-specific Criminal Justice terminology. The primary objective was to determine if particular given names had a higher rate of being subject to systemic racial bias in the criminal justice system than others. I utilized HAL to replicate human stereotypes, while creating a 1-7 Likert Scale to exemplify natural human judgment of given names and their potential criminality.

(2) *Exploratory Text Analysis with R and Shiny*

David Allbritton, DePaul University

Text analysis and text mining are powerful tools for automatically extracting patterns from unstructured texts, such as texts from open-ended survey questions in program evaluation or research. The open-source statistical programming language R includes powerful modules for text analysis such as tm (Text Mining) and NLP/openNLP (Natural Language Processing) (http://cran.r-
Shiny (http://shiny.rstudio.com/) is a web application framework for creating interactive web interfaces to data analysis programs in R. Analysis outputs are wrapped in html markup for viewing in a web browser. Input forms allow the user to change analysis parameters interactively and immediately view the results. A Shiny application is presented for exploratory analysis of open-ended text data, such as course evaluation comments. Texts are analyzed based on lexical n-gram frequency and grouped into “topics” using Latent Dirichlet Allocation (LDA). The user can change analysis parameters (such as minimum frequency, whether to use stemming, n-gram size, number of topics, etc.) interactively while viewing the results. This tool could be useful for generating coding categories prior to doing a traditional content analysis, or for automated exploration of a large set of open-ended responses. Source code is provided for use as a teaching tool or further development.

(3) WhatsOb: An Android app for note-taking during behavioral observations

Julia Watzek, Georgia State University
Sarah F. Brosnan, Georgia State University

For behavioral observations of both humans and animals, it is crucial not to lose focus of the subjects at hand, and yet take detailed, accurate, and timely notes. Mobile devices are highly versatile, unobtrusive, and portable tools that lend themselves to the task. While there are numerous note-taking apps, they typically have several shortcomings for use as scientific tools: 1) typing on a touchscreen keyboard is tedious, takes time, and is prone to typos; 2) dates and times have to be recorded manually; 3) the raw data typically require extensive reformatting prior to data analysis. To address these problems, we developed WhatsOb, an Android app designed to allow for fast but accurate data capture with minimal device interaction. Preconfigured quick text buttons replace tedious typing and researchers can set up flexible observation formats for different study populations. The app allows configuration of several sets of quick text buttons to minimize screen clutter and increase versatility. Additionally, all entries are time-stamped automatically and researchers can schedule timers in the form of sound or onscreen notifications. Data files are exported as plain text or comma-separated values (CSV) files via email and common cloud file hosting services. We demonstrate usage of the app for three types of behavioral observations: scan samples, focal samples, and ad lib data collection.

(4) Evaluating Equating Methods for a Manual and Computerized Card Sorting Task

Christopher Koch, George Fox University

Different versions of the same test are typically equated so that scores can be similarly interpreted over time. Likewise, different versions of the same test may need to be equated. For instance, the Nonverbal Stroop Card Sorting Task (NSCST; Koch and Roid, 2012) is a manual card sorting task. Although computerized version of the task simulate the NSCST, the response time results are different (Koch and Hottovc, 2012). Two of the most frequently accepted methods for equating are linear and equipercentile equating. These two methods of equating are compared for the NSCST using a sample (n = 30) of middle schoolers. Administration order of the NSCST and the computer tablet version of the test was randomized across participants. The two tests were equated based on raw scores and verified.
using standardized scores. Overall, the equipercentile method produced the best fit. Implications for equating performance measures are discussed.

(5) Attention Motivation through Gamification

J. Antonio Salamanca, Georgia State University
David A. Washburn, Georgia State University

Executive attention involves the willful control of processing, and assumes that individuals have both the ability and the motivation to inhibit prepotent response tendencies, to ignore salient but irrelevant sources of distraction, and to focus on and learn in contexts that might be boring. In this presentation, we review multiple studies in which we attempted to manipulate motivation by introducing game-like elements to simple learning or performance tasks. Two categories of manipulation will be reported. First, we compared learning and memory for content that was delivered to participants in an online format, but in which access to the online material either required simple mouse-click or required solving game-like puzzles (word-search, anagrams, matching-game). Initial learning and subsequent retention were both significantly improved for the material encountered in the game interface, even though it was actually easier and more efficient to access that information in the absence of the game. Second, we examined performance on tasks in which the visual appearance and response characteristics were identical for the two groups of participants; however, one group was given a game-like cover story whereas the other group’s instructions simply described the stimuli as they appeared on the screen. Again, performance differences were observed, suggesting that motivation was engaged by the gamification of the assessment--even thought this did not alter the appearance of the task at all! Limits to performance changes from gamification were also documented, underscoring the fact that individual differences in motivated attention involve both between-individual differences and within-individual variability.

(6) Smart Therma-glove System for Raynaud's Disease

Michael Carando, Wentworth Biomed
Damon Cole, Wentworth Biomed
Nicholas Poplar, Wentworth Biomed
Weihui Li, Wentworth BME
Chen-Hsiang Yu, Wentworth

Raynaud’s disease is the constriction of arterioles in the fingers and toes. Symptoms include cold fingers/toes, leading to white or blue skin color and numbness. These symptoms are triggered by emotional stress and cold weather conditions. Heated clothing has commonly been used to help treat Raynaud’s sufferers by keeping extremities warm. Using existing heated apparel as a base, this study built a self-regulating, smart heated glove to treat and prevent Raynaud’s symptoms automatically. The smart glove features various biosensors to collect pulse rate, hand skin temperature, and local environmental temperature, all integrated into the Arduino UNO board. Pulse rate is measured through a pulse sensor or photoplethysmography mechanism. These readings are then used to measure stress levels through heart rate variability analysis. Hand skin temperature is monitored by a thermistor (MCP9700). Local environmental temperature is tracked through an android smart phone “weather” app and transmitted to the microcontroller with Bluetooth Low Energy communication.
(nRF8001). Based on the collected parameters, the microcontroller can automatically turn on or off two heating pads (Sparkfun COM-11289), and regulate its temperature level as needed with pulse width modulation. The system is powered by a rechargeable external battery pack to allow for extended use to combat stressful. The Smart Therma-glove gives Raynaud’s disease sufferers a convenient way to manage symptoms automatically.

(7) Using Interactive 3D Software to Create Manipulatable Human Figures for Body Perception Research

Grayson Lein, Linfield College
Kay Livesay, Linfield College

The paper presents the use of the DAZ3D program as a measurement tool for body size perception. When studying body schema, researchers often rely on human figure comparisons to examine body size perceptions. Often these figures are two-dimensional drawings or photos of human bodies. However, human bodies are three-dimensional. Previous research has shown the advantage of using three-dimensional changeable figures in assessing body size perception (Crossley, Cornelissen, & Tovee, 2012). We chose the DAZ3D program over other options (e.g. Body Visualizer) because it allows the user to rotate the figure in space (both depth and plane), convert manipulated figure measures to real life metrics (e.g., inches or centimeters), input real life metrics to create figures, and manipulate over 50 parameters of measurement consisting of both length and circumference. The downside to DAZ3D is that it can be confusing to set-up and use. We explain how to use DAZ3D software effectively for use in body size perception research. We had participants use the DAZ3D software to represent their own body, allowing them to manipulate 17 body measurements. Our data suggests that participants can easily use the program and accurately represent their body size (their figure was compared to real life body measurements). Additionally, because DAZ3D has the ability to manipulate almost all aspects of the human figure (including parameters such as muscle mass), researchers will be able to make a more fine-grained analysis of distortions in body perception in both men and women.

(8) Using dynamical models to test experimental hypotheses in repeated-measure designs

Henry S. Harrison, University of Connecticut, henry.harrison@uconn.edu

Dynamical systems theory provides a useful language for describing systems of interest, but the use of dynamical models in behavioral research is limited by practical difficulties such as the necessity of fitting parameters to observed data. Brute-force methods are computationally expensive, and for chaotic systems general-purpose optimization techniques are unlikely to produce satisfactory solutions due to instability of the model-data manifold. A method called dynamical parameter estimation (DPE) addresses these issues by introducing a coupling term that synchronizes the model to the observed trajectory during the fitting process. Thus, the problem becomes one of minimizing the necessary coupling rather than minimizing the difference between the observed and simulated trajectories. DPE stabilizes the model-data manifold, allowing the use of general-purpose optimization algorithms such as sequential least squares. However, further consideration is required when multiple trajectories are observed. In an experimental setting, observed trajectories arise under different values of explanatory variables (i.e., fixed effects), often for different units of observation (i.e., random effects). Thus, model parameters may be expected to vary systematically across observed trajectories. Here, I present a
multilevel approach to parameter estimation that embeds a mixed-effect regression within DPE. This allows for the testing of experimental hypotheses as part of the parameter-fitting process. Simulation studies confirm that systematic variations in model parameters can be recovered as expected. An implementation of the method is published alongside this presentation under an open-source license. This software removes barriers to the inclusion of chaotic dynamical models within an experimental research paradigm.

(9) Minding the Details: Use of Mind-Mapping Software to Teach Psychology and Probability

Will Whitham, Georgia State University
David Washburn, Georgia State University

Tools used to create attractive visualizations of hierarchies of lists, often called “mind-mapping” software, are lightweight, intuitive, and have been advertised for use team-based brainstorming, organizing hierarchical information, and creating study tools. Such concept maps have demonstrated some effectiveness for the teaching of various subjects (Nesbit & Adesope, 2006), and the virtues of outlining and sketching more generally are well-documented (Tversky, 2002). We explored the feasibility of using such software to teach introductory lessons in psychology and probability. Specifically, undergraduate students spent approximately an hour with either a computerized lesson on the history of psychology or a lesson on basic problems in probability theory. This lesson was either presented in a more conventional format or with the addition of mind-mapping visualizations. We report the relative retention of information by students in each group, and discuss other ways in which such software can be applied in the classroom.

(10) Narrowing in on the Timing of Semantic Representation in the Brain using Multiband fMRI during Reading

Benjamin Schloss, The Pennsylvania State University
Ping Li, The Pennsylvania State University

An important assumption of computational models of word meaning like distributed semantic space models is that words occur together in a non-random fashion, such that the meaning of a particular word may be inferred from the company that it keeps. Encouraging results have emerged suggesting that these co-occurrences can predict patterns of brain activity from functional magnetic resonance imaging (fMRI) data (Mitchell et al, 2008) for a large variety of words (Mason & Just, 2016) in varied contexts such as listening to fluent speech (Huth et al, 2016) and reading (Wehbe et al, 2014). However, most fMRI experiments acquire one full brain image every two seconds—enough time to read three words or an entire sentence depending on the paradigm. Consequently, researchers have resorted to averaging all brain images corresponding to a target word (Yarkoni et al, 2008), assuming that the words appearing in a two second time window around a particular target word are random, contradicting the assumptions of the computational models used to study and decode the same patterns of brain activity. To date, no practical assessment of the effects of temporal resolution and averaging on decoding activity patterns has been conducted. In the current study, we analyze 10 subjects’ data in a self-paced reading experiment which employs multiband imaging methods to achieve a temporal resolution of one full brain image every 600ms and compare the effects of
averaging representations for single words at different temporal resolutions on the decoding accuracy of semantic space models.

(11) Continuous Response Interface Boosts Online Donations

Tianwei Liu, Meritco Services
Jing Chen, New Mexico State University

Raising charitable donations requires tremendous time and effort. Online giving accounts for an increasing share of funds raised by non-profits, and web design is critical. Despite the large variations in design, most websites display donation amounts with discrete options (e.g., $10, $20, $50). We argue that this discrete interface may cost some potential donations because when a donor’s preference falls between two discrete options she is likely to choose the lower one. In light of this reasoning, three online experiments (total N = 1364) examined whether a continuous input interface, presented as a slider bar, may bring more donations than the discrete interface. In Experiment 1, participants were randomly assigned to a discrete or continuous interface and made hypothetical donations to a local radio station. Those in the continuous condition were more likely to donate (i.e., higher donation likelihood) and made larger donations (i.e., greater donation amount). Experiment 2 used real donations. Participants were given $2 and then asked to donate to UNICEF. Likely due to a small change in the experimental design, the donation likelihood pattern failed to replicate, but the donation amount pattern still holds. Experiment 3 included “donate zero” as a standalone option to make the scenario as close to real-life donations as possible, using hypothetical donations. Again we replicated the donation amount pattern but not the donation likelihood pattern. The implication of these results is that with proper design, continuous interface may be a better choice for eliciting greater donation amounts than discrete interface.

(12) The Political Semantic Space from 2000 to Present

Sha Liu, The Pennsylvania State University
Benjamin Schloss, The Pennsylvania State University
Ping Li, The Pennsylvania State University

In this study, we analyze political concepts from American presidential debates from 1999 to 2016 using a computational model of word meaning. We compare the representations of political concepts between the Republican and Democratic parties based on the debates from 2000, 2008 and 2016, because they are big election years. In our analysis, all the debates comprise a single semantic space. However, a set of hand chosen, politically charged words are tagged based on the speaker as Republican or Democrat, and by the election year. Thus, we are able to compare different representations of the same word between the parties and over time. We also compare the conceptual organization of political concepts from the 2016 election for Hillary Clinton and Donald Trump. Our results indicate both intuitive behavior of the semantic spaces as well as some alignment with popular opinions about the parties and current election. For example, we find that concepts derived from a single individuals’ speech (Hillary and Donald) are more cohesive than concepts derived from the debates of an entire party. As for the two parties, our results indicate that their political semantics have diverged since the 2000 election; however, these divergences is not simply linear over time. As for Clinton and Trump, Clinton’s derived political semantics are closer to that of the average
Democrat than Trump’s are to the average Republican. Finally, we report a detailed graph based analysis on the centrality of different political terms in the different parties over the three big election years.

(13) Ti-Med

Nicole McColgan, Wentworth Institute of Technology
Benjamin Ford, Wentworth Institute of Technology

As healthcare technology continues to grow rapidly, more potential prescription medication will be available to patients with ailments. In elderly patients, who may have multiple health related problems, it is not uncommon to have many prescription medications that need to be taken in synchronization. The problem arises when these elderly patients forget to take their medication on time, refuse to take them, overdose, or take the wrong prescription. This can lead to adverse side effects that will further complicate their already existing medical condition. Not only does this negatively affect the patient directly, it also causes hardship for the family as well as difficulty for the clinician to effectively treat the patient. The study designed a medication dispenser that elderly patients can use to ensure they are taking the correct dosage of medication at the correct time. This “pillbox” is built on a Simblee microcontroller system with low energy Bluetooth module integrated. A force sensor is used to detect the open action of the pill box. It will trigger an alarm to alert the patient that it is time to take their medication if force sensor has not detected the corresponding pill box being opened on time. Additionally, notifications are sent to a mobile device app for a family member or caretaker for confirmation.

(14) Peltier Effect Driven Fluidic Cooling Vest for Multiple Sclerosis

Alyssa Payette, Wentworth Institute of Technology
Ahmad Ateyat, Wentworth Institute of Technology
Weihui Li, PhD, Wentworth Institute of Technology

Multiple Sclerosis (MS) is a disease of the central nervous system that has many debilitating symptoms. Among those, is a high sensitivity to heat, which adversely affects those suffering from MS as it impairs the ability of the demyelinated nerve to conduct electrical impulses. Various cooling therapies have been developed to counteract the detrimental effects of heat on patients suffering from increased body temperature such as cooling clothing garments. The study developed an automatically thermoregulated cooling vest system. This system uses the Peltier effect to cool water in a reservoir located within a small backpack. This chilled water is then pumped through tubing that is embedded into the person’s vest. A temperature sensor located in the underarm area of the vest reads the temperature of the person. When the temperature exceeds the threshold, it triggers the system to run until the temperature returns to a value below the set temperature. The backpack system has an easy connection to the vest and the vest itself is composed of a lightweight, breathable material that is easy for the user to put on. The whole system is achieved through an Arduino system which collects temperature data, controls the water pump and drive the Peltier generators and support fans. The cooling vest system will help MS patients to better manage their body temperature and improve their quality of life.
(15) Bluetooth Insulin Pump System to Reduce Social Stresses of Insulin Pump Use

Elizabeth Suitor, Wentworth Institute of Technology
Mary Rodgers, Wentworth Institute of Technology
Caitlin Corcoran, Wentworth Institute of Technology
Weihui Li, Wentworth Institute of Technology

Type 1 Diabetes affects one in three hundred people by the age of eighteen in the United States. While insulin pumps are effective in the treatment of Type 1 Diabetes, many patients choose not to use the devices due to the fact that their use often means a breach of personal privacy in public as well as the inconvenience of estimating the dosage, which can lead to mis-dosage. The goal of this project is to develop a Bluetooth enabled wearable wristwatch and mobile application that directly communicates with the pump to administer insulin. The watch is based on Simblee microcontroller system with integrated low energy Bluetooth. It has the ability to calculate dosage, control the pump for insulin administration and display time. The mobile application includes all of these features as well as health data logging so the effects of the insulin on their system can be better understood and an alarm system where patients can program the watch to ring at certain times, giving them reminders that they need to take their insulin. This watch allows easy, convenient diabetes management allowing the user to privately administer insulin from the pump. In doing this, it can reduce the number of missed or delayed doses. This system will reduce much of the anxiety related to using an insulin pump while also increasing the effectiveness with which insulin pump users treat their diabetes.

(16) Careful what you post: Pronoun choice exacerbates emotions in tweets about terrorism

Eliza Barach, The University at Albany SUNY,
Samira Shaikh, University of North Carolina Charlotte
Yousri Marzouki, University of Aix Marseille
Laurie Beth Feldman, The University at Albany SUNY

The current study compares about 40,000 English tweets produced about the Charlie Hebdo shootings in February 2015 and about 40,000 English tweets produced in reaction to the November 2015 Paris terrorist attacks. All tweets had a hashtag that included terms such as Paris, terrorism, or attack, associating it with the terror events in Paris at that time. Tweets were then broken down into their constituent words and how many times each occurred. Analyses were conducted separately for tweets with each pronoun and were based on key words with extreme (15%) negative or positive polarity based on Warriner et al., 2013. Valence and arousal of key words (frequency weighted and logged) were significantly more extreme for tweets with variants of the “I” or “we” pronouns than for tweets without personal pronouns (baselines). Negative emotions (valence and arousal) in tweets without pronouns, associated with the November Paris attacks were more extreme than earlier Charlie Hebdo events. With respect to group dynamics and psychological states (Chung and Pennebaker, 2007), differences between I and we were more reliable for words with a negative polarity. Key words in February tweets in “I” tended to be lower in arousal than those in tweets in “we”. November tweets in “I” tended to be more negatively valenced and higher in arousal than tweets in “we”. Whether the absence of reliable collective versus individualists’ differences for words with positive polarity is reliable is under further exploration.
**(17) Improved EDA Measurement Device**

Zachary Schneider, Wentworth Institute of Technology  
Kurush Mirza, Wentworth Institute of Technology  
Ismail Alzahrani, Wentworth Institute of Technology  
Weihui Li, Wentworth Institute of Technology

Electrodermal activity (EDA) constitutes a non-subjective psychophysiological metric useful for detecting changes in stress, cognition, attention, and emotion. In particular, EDA measurement devices generate clinically relevant data that is commonly employed to assist healthcare practitioners in monitoring patients of a range of neurological and psychiatric disorders, such as autism and epilepsy. Clinical research has established relationships between emotional arousal, sympathetic nervous system activity, and electrical potential across areas of human skin. The present study is to build an EDA measurement device that can continuously monitor stress level of the users. The device uses a Simblee BLE model to control the delivery of milli-ampere level currents to the body through two electrodes and measures the changes of skin conductance. It then transmits the real-time data to a program written with Processing to display the data. The data are also processed through fast Fourier Transform to find the stress level (the relative magnitude of sympathetic and parasympathetic never activities). An Android application is also designed in order to display this signal data trace on a mobile platform, as well as to incorporate software-based signal analysis functionality such as derivative calculation and FFT and therapeutic functions such as medication remind and guidance.

**(18) User's Experience and Strategy Use with Pokémon Go**

Elizabeth M. Cedillos-Whynott, Miami University  
Hayley Skulborstad, Mt. Ida College  
Aimee Belanger, Miami University

The new mobile app, Pokémon Go, was released this past summer and has quickly become the most downloaded app of all time. The app is an augmented reality, location-based mobile game based on the popular Nintendo video game, Pokémon. Similar to the video game, where the player explores a virtual world in order to capture and train creatures called Pokémon, the app requires the player to explore the real world and will populate the user’s world with virtual Pokémon. The vast engagement with the app has led to many unexpected outcomes, both positive and negative. On one hand people are becoming more physically active, as the game requires exploring new areas and surroundings in our physical world. Further, they are becoming more social, as there are frequent encounters with other players while capturing Pokémon. However, some have criticized the game as removing a person from reality, causing distractions that lead to accidents because users are not paying attention to their surroundings. It has also been criticized as causing obsessive behavior and being overly time consuming. A survey will be distributed to introductory psychology undergraduates. The survey will include questions about user’s experience playing the game, including personal and social enjoyment, user’s strategy while playing, and their decision making to further their game experience. Preliminary data regarding how players engage with the app, the social elements involved in the game, how players make judgements and decisions in the game and their motivation for these actions will be discussed.
(19) Struggling Adult Readers Spend More Time Choosing to Learn, but Achieve Comprehension Strategy Learning Comparable to Normal Readers

Breya Walker, The University of Memphis
Raven Davis, The University of Memphis
Whitney Baer, The University of Memphis
Arthur Graesser, The University of Memphis

The present research investigates user performance during AutoTutor training of reading comprehension strategies developed in the context of the Center for the Study of Adult Literacy (CSAL). In a between-subject design, college students (N=34) and struggling adult readers (N=45) interacted with one review module that assesses multiple reading comprehension skills (i.e., making inferences, acquiring vocabulary, clarifying and evaluating texts). The design of the module was similar to Jeopardy, with questions (N=16) being categorized as either Easy or Hard dependent upon their placement on the board and the skills required to answer them. We hypothesized that struggling readers would spend more time on the review lesson compared to college students, they would opt for easier questions given their reading ability, and their performance would be different compared to college students. Our results reveal that there was no difference in the amount of time struggling adult readers spent answering Easy versus Hard questions. There also were no significant group differences in which question types (i.e., Easy vs. Hard) were selected to answer. However, struggling adult readers spent more time selecting and answering questions, regardless of their question type, compared to college students. These data suggest that struggling adult readers spend more time interacting with the AutoTutor training module regardless of question type but perform just as well as college students in answering the questions.

(20) ECG Drowsiness Detection System

David A. Puskar, Wentworth Institute of Technology
Angely Rodriguez, Wentworth Institute of Technology
Weihui Li, Wentworth Institute of Technology

Drowsiness impairs a driver’s ability to drive as the body’s responses such as mental agility shut down gradually. To target the drowsiness problem during driving, various alert systems have been produced and marketed over the years. Nevertheless, many of the systems are either fairly expensive, such as facial recognition systems or they simply do not work efficiently, such as the head-tilt detection devices. Research found that when people get drowsy, their heart rate declines. The present study designed a driving alert system by tracking the driver’s heart rate. To build the driving alert system, three dry metal electrodes are placed on the steering wheel. When driving, two palms and one right-hand thumb rest on these metal electrodes to collect the raw ECG. The signal is then filtered and amplified by an AD8232 chip. It is further digitalized and analyzed by Arduino UNO microcontroller. During the analysis, half-second long arrays are collected and de-trended. A confirmation of whether or not there is a peak of the QRS wave is used to determine the heart beat and calculate heart rate. Once a reduced heart rate is detected, an alarm will sound off. In addition to this, should the driver doze off and their hands slip from the wheel, an alarm will also sound. In conclusion, this device can monitor a user’s ECG while their hands are on the wheel and detect a decrease in heart rate related to drowsiness.
(21) Force Sensing Insoles with Therapeutic Replacement Padding

Matthew Caron, Wentworth Institute of Technology
Benny Wu, Wentworth Institute of Technology
Hung Tran, Wentworth Institute of Technology
Weihui Li, Wentworth Institute of Technology

Bio-mechanically the foot is one of the most complex structures in the human body’s skeletal system. The average sedentary person may take about 1,000 to 3,000 steps a day whereas an athletic person may take over 10,000 steps a day. The different types of lifestyles leave many people suffering from foot pain which may cause other pains in the body as well. Pain in the feet can cause individuals to change their strides and walking patterns according to the burden of discomfort in their feet. The present study designed an insole that utilizes sensors to record the ground reaction forces, which can be used to reconfigure the insole conveniently by changing interchangeable padding. The insole integrates six sensors (FlexiForce A401) to monitor the ground reaction forces on the heel and ball of your foot near the toes. These sensors can help doctors to detect the greatest source of pressure during a patient’s gait cycle. This data will further allow them to detect irregularities in their gait patterns that may cause pain or joint disfigurement. These sensors wirelessly relay the data to a computer through an embedded micro-system. The insole has interchangeable padding allowing the user to alter its comfort and support in specific areas depending on origin of pain and data readings from the sensor.

(22) Web-based Experiment Management

Joerg Beringer, BeriSoft LLC

Traditionally computerized experiment software solutions focused on the implementation of the test paradigm, either with providing an authoring environment or the implemented test itself. With web-based research, this focus now expands from availability of test paradigms to the management of the experiment in the cloud as a service. Web-based research typically is associated with the recruiting of large populations of anonymous volunteers, but the cloud offers other opportunities that are interesting for anyone who wants to streamline their lab operations and take advantage of modern web technology. This presentation will give an overview about advanced setups which you only can do if your experiment is managed in the cloud: Multi-channel test applications: Using open-source libraries such as jsPsych or programming from scratch you can write HTML5/Javascript test applications that run on multiple operating systems and devices including mobile in a web-browser that provides sufficient runtime accuracy. Subject Enrollment: While web-recruiting like Mechanical Turk is maybe the most extreme use of web-based data collection, the ability to plan and schedule sessions in form of a web service opens up new scenarios such as inviting subjects per email or test message notification or posting experiments for self-registration on intranet. Multiple deployment scenarios: Hosting an experiment in the web, even facilitates the data collection in well-controlled lab environment when using multiple test stations. In addition, it gives you the possibility to collect data in multiple lab locations, in the field, or from home with the same test infrastructure. Single source of truth: With all data being stored at one database, you can monitor progress of session as well as review performance...
as soon as a session has been completed. This increases the quality of data and simplifies radically data management when closing an experiment.

(23) Complexity Matching and Linguistic Convergence Across Bilingual Interlocutors

Sara Schneider, University of California, Merced
Adolfo Ramirez-Aristizabal, University of California, Merced
Alexis Luna, University of California, Merced
Gilbert Sepulveda, University of California, Merced
Christopher Kello, University of California, Merced

Many studies have shown that, when people converse, various aspects of their speech and language patterns often converge to some degree. Recent work on complexity matching discovered a new type of convergence in the statistical clustering of acoustic events in conversational speech signals. Results to date indicate that clustering reflects speaking style rather than the specific words and sentences spoken, which suggests that complexity matching can occur even when conversing dyads speak different languages. To test this prediction, 28 pairs of Spanish-English bilingual speakers engaged in each of three different conversations: all English, all Spanish, and a mixed condition in which one participant spoke only English while the other only spoke Spanish. The results displayed strong evidence for complexity matching in all three conditions on timescales starting around 300-500 milliseconds and longer, i.e. timescales for which speakers can adjust their speaking styles. The results showed that complexity matching spanned across different conversations, demonstrating that convergence in speaking style does not reflect any local measure of alignment, nor does it reflect the particular words and sentences being spoken. These findings demonstrate a new and powerful method for measuring convergence in the temporal structures of conversations.

(24) Towards Understanding Lexical Psychological Phenomenon and Mental Engagement via the Stochastic Model of Buddhism Psychology

Shivani Poddar, IIIT Hyderabad

For the past several decades only a handful of models attempt to capture the lexical persona and exhibited mental states of an individual. These models include, but are not limited to the Big5, MBTI models of personality. By means of this work, we attempt to elucidate the Buddhist Model of Persona so as to address and counter the staticity of these models. This model leverages from the Abhidhamma traditions of psychology and is constructed using a stochastic state machine to capture the moment by moment varying lexical psychology (and the concomitant mental states) of an individual. We especially draw focus towards the various dynamically evolving psychological phenomenon (here Anxiety and Mindfulness) that can be observed, inferred, and recognized as evinced by a social media user through this model. Justifying the defining properties of Abhidhamma which are foundationally more fitted to undertake this study we analyze it in context of theoretical psychology (for instance Brentano’s ideas) and phenomenology. Finally, we propose a lexical mental engagement factor, which helps us in quantifying the engagement illustrated by an individual towards a given psychological phenomenon or mental state. We conclude by discussions pertaining to various personality oriented classifications of individuals based on this engagement factor. Our inferences align well with the doctrines of western
psychology and biomedical explanations of the phenomenon covered, thus validating the theories proposed suggested by our model.

(25) Evaluation of Dialogues and Triialogues in ITS

Keith Shubeck, The University of Memphis
Ying Fang, The University of Memphis
Vallabhaneni Sree Lasya, The University of Memphis
Xiangen Hu, The University of Memphis

Math comprehension and learning has been found to be enhanced with the presence of self-reflection and deep level reasoning questions, even in simple vicarious learning environments with animated agents. Algebra I mathematics lends itself to the use of stepwise worked examples. This stepwise problem setting allows for a large amount of interactions for the human learner with animated agent(s), as well as opportunities to inject self-reflection and deep-level reasoning questions into dialogues and triialogues. Conversation-based learning environments provide opportunities for intelligent tutoring systems (ITSs) to closely model human tutoring interactions. The experiment described in this paper explores and tests the added effectiveness of interactive dialogues and triialogues in learning Algebra I, utilized in a modular, online-based ITS which takes advantage of a SKO (shareable knowledge object) framework and natural language processing. The experiment compares learning across five conditions: (1) a static reading control condition, (2) a vicarious control dialogue condition with animated agents, (3) an interactive dialogue condition (i.e., human learner and tutor agent), (4) an interactive triialogue condition (i.e., human learner, tutor agent, and tutee agent) and (5) a vicarious monologue condition. Triialogues in ITSs have been shown to promote learning in a variety of domains (e.g., scientific reasoning, reading comprehension). This paper will seek to answer questions concerning the effectiveness of dialogue and triialogue conversation environments in an Algebra 1 domain compared to vicarious learning, and whether or not triialogues provide an added benefit over dialogues within this domain.

(26) AXIS: Generating Explanations at Scale with Learner sourcing and Machine Learning

Joseph Jay Williams, Harvard University
Juho Kim, Stanford University & KAIST
Anna Rafferty, Carleton College
Samuel Moldonado, WPI
Krzysztof Z. Gajos, Harvard University
Walter S. Lasecki, University of Michigan
Neil Heffernan, WPI

While explanations may help people learn by providing information about why an answer is correct, many problems on online platforms lack high-quality explanations. This paper presents AXIS (Adaptive eXplanation Improvement System), a system for obtaining explanations. AXIS asks learners to generate, revise, and evaluate explanations as they solve a problem, and then uses machine learning to dynamically determine which explanation to present to a future learner, based on previous learners’ collective input. Results from a case study deployment and a randomized experiment demonstrate that AXIS elicits and identifies explanations that learners find helpful. Providing explanations from AXIS also
obj ectively enhanced learning, when compared to the default practice where learners solved problems and received answers without explanations. The rated quality and learning benefit of AXIS explanations did not differ from explanations generated by an experienced instructor.

(27) Learning from errors by doing in online math tutoring system

Jun Xie, The University of Memphis
Scotty Craig, The University of Memphis
Art Graesser, The University of Memphis
Xiangen Hu, The University of Memphis

This study examined how 6th graders learned math from errors while working with an educational technology program, ALEKS (Assessment and LEarning in Knowledge Spaces). The learner’s strategies after an error were categorized as explanation (students only read explanations in the next two attempts after an error), practice (students only solved problems in the next two attempts after an error), and mixed (students read an explanation and solved a problem in the next two attempts after an error). An attempt meant that a student read an explanation or solved a problem. An error was an attempt when a student failed to solve a problem. Results indicated that the percentages of practice (M = .31, SD = .16) and the mixed strategy (M = .33, SD = .18) were significantly higher than explanation (M = .27, SD = .15), F(2, 2854) = 731.06, p < .000, η2 = .69. Also, when compared to explanation (M=.25, SD = .20), both practice (M = .36, SD = .13) and mixed strategy (M = .37, SD = .12) led to significantly higher probability of the correctness on the next attempt, F(2, 964) = 64.72, p < .000, η2 = .12. Further analysis of the mixed category implied that when students read an explanation before solving a problem (M = .38, SD = .14), it caused significantly a higher probability of the correctness on the next attempt than solving a problem before reading an explanation (M = .36, SD = .14), t(322)= 2.44, p<.000, d = .07. So students would be more effect by doing, especially doing after an immediate explanation, when learning from errors with ALEKS.

(28) A Testing Utility for AutoTutor

Qizhi Qiu, The University of Memphis
Zhiqiang Cai, The University of Memphis
Xiangen Hu, The University of Memphis
Arthur Graesser, The University of Memphis

AutoTutor is a framework that adds deep reasoning conversations to intelligent tutoring systems. The AutoTutor Script Authoring Tool (ASAT) allows domain experts, language experts, media developers and programmers to collaboratively create high quality conversational tutoring lessons. An AutoTutor script is a package containing interactive media pages, agent speeches, questions, answers and interaction rules. However, the current ASAT provides limited testing functionality. This paper proposes a testing utility that provides complete and systematic support to authors and independent testers for testing every authored element throughout the authoring process. AutoTutor script testing tasks include agent speech testing, media interaction testing, semantic match testing, conversation rule testing and assembled lesson testing. Although the testing utility focuses on system testing, it is equipped with tools that employ widely used software testing techniques including black box and white box testing methods. For example, equivalence partitioning is used for basic functionality testing,
DD-path (Decision-to-Decision path) testing of conversation rules and web testing for media interaction testing. Test case information is stored in a database that is used for further regression testing and bug reproduction. In addition, texts supplied by the testers can be assimilated into the scripts; for example, a tester’s answer to an AutoTutor question could be used by the student agent. While human interaction during the testing is unavoidable, the utility automates testing whenever possible. This fully integrated testing utility makes it possible to test AutoTutor scripts at authoring time and thus improve both the testing efficiency and the quality of AutoTutor lessons.

(29) Topic Network Analysis on a Summarization Corpus

Yan Gong, The University of Memphis
Zhiqiang Cai, The University of Memphis
Haiying Li, The University of Memphis
Xiangen Hu, The University of Memphis
Arthur Graesser, The University of Memphis

Topic modeling discovers the topics in a collection of documents. Once the topic model is built, a document can be represented by topic scores, indicating how much or what proportion of a specific topic is contained in the document. However, the topic scores do not give information about how the topics are associated within each document. It has been shown in Epistemic Network Analysis (ENA) that such associations are important in language analysis. In this paper, we use ENA tool on top of topic models to explore how much the topic association helps in understanding documents. We used a corpus collected from a summarization study, containing 1481 student summaries on eight (8) target texts. The task is to correctly classify the summaries into 8 groups based on their target texts. Our earlier study showed that topic proportion score was unable to perform this task very well. In this study, we want to see if the topic association helps. We split each summary into sentences and computed topic scores for each sentence. Then we computed the association strength between topics by multiplying the corresponding topic scores. The sentence association scores were added up to form the association score of a document. Principle component analysis were then applied to project the document association scores into a low dimensional space. The low dimensional vectors were used as features for summary classification. The results showed that the association scores were more powerful than the raw topic scores in this task.

(30) Data collection and sharing in AutoTutor System using Experience API (xAPI)

Qinyu Cheng, The University of Memphis
Xiangen Hu, The University of Memphis
Zhiqiang Cai, The University of Memphis
Arthur Graesser, The University of Memphis

AutoTutor is an intelligent tutoring system that simulates a virtual real world classroom by holding dialogs (human with one agent), trialogs (human with two agents), or conversation with even more agents with human learners in natural language. AutoTutor is adaptive to the learners' actions, verbal contributions, and in some systems even their emotions. The learning experience including student behavior performance and learning progress data in AutoTutor is very complicated due to its adaptability. Before using Experience API (xAPI), big data analysis was very time consuming, and even
then that the data was not easily tracked and shared. The xAPI specification released by Advanced Distributed Learning provides a solution for this problem. It describes learning experience with activity streams and records experience as statements. The xAPI specification provides a standard solution to transform the unstructured data into structured data that are easily expandable by defining the semantic meaning. In order to help the researcher identifying conversation patterns in a systematic way, and it is imperative to store data in data streams by using xAPI. The advantage of applying such an action stream has two big benefits. First, it saves time and storage by processing data once; second both cognitive scientist and psychology scientists can easily and directly track and share student’s behavior performance and learning progress to do the analysis due to the standardization (actor-verb-object). The efficiency and effectiveness gained in the rich data analytics and visualization of education system is a big win for every partner.

(31) Conversation-based Assessment Systems created with AutoTutor Script Authoring Tool for Assessment

Carol M. Forsyth, Educational Testing Service
G. Tanner Jackson, Educational Testing Service
Diego Zapata-Rivera, Educational Testing Service
Blair Lehman, Educational Testing Service
Irvin Katz, Educational Testing Service

Conversation-based Assessment systems (CBAs; Jackson & Zapata-Rivera, 2015; Zapata-Rivera, Lehman, & Jackson, 2015) have been created at ETS to assess students’ knowledge and skills in domains including science, math, English language arts (ELA), and argumentation. The purpose of creating these systems is to assess students’ skills based on information that they may be able to articulate in natural language conversations, but may not be easy to assess using static item types (e.g., multiple-choice questions). CBAs were created with the AutoTutor Script Authoring Tool for Assessment that allows authors to augment the tutorial conversational framework of AutoTutor for assessment purposes and provide natural language capabilities of the original AutoTutor (see Graesser, 2016). With this tool, we have created and tested multiple dialogic structures, altered versions of the original tutorial structure, to probe students to explain their understanding rather than transmit new knowledge. In addition to dialogic changes, CBA systems have implemented the latest technology including speech recognition for assessing English-language learners and interactive interfaces to assess listening comprehension and science inquiry skills. Furthermore, complex conversations and visual displays make it possible for CBA systems to help properly attribute students’ performance to their strengths and weaknesses. For example, one CBA system disentangles ELA from mathematics skills, diagnosing when students might not have the necessary vocabulary to express the answer to a mathematics word problem. At the conference, we will demonstrate this and other CBA systems that were designed to assess a wide array of skills through natural language conversations and interactions.

(32) Learning from OperationARA: A Game-like version of AutoTutor

Carol M. Forsyth, Educational Testing Service
Keith Millis, Northern Illinois University
Zhiqiang Cai, The University of Memphis
Arthur Graesser, The University of Memphis
OperationARA (the online version of OperationARIES!; Millis et al., 2011) teaches students research methodology by extending the AutoTutor framework (for a review, Graesser et al., 2012) to allow students to have natural language conversations with two or more artificial agents, referred to as trialogues. These trialogues allow for additional affordances from the original single agent model in AutoTutor (Graesser, Forsyth, & Lehman, in press). Specifically, the two agents make it possible for the system to dynamically adapt to students varying levels of prior knowledge. For example, a student with a low-level of prior knowledge may benefit more from simply watching a teacher agent teach a student agent whereas a student with high prior-knowledge may benefit more from teaching a student agent. Other pedagogical features include multiple-choice questions, an E-Text, and hint lists. Beyond these pedagogical features, there are also game-like features including a narrative about space invaders who are publishing bad science. The students must learn scientific inquiry skills to save the world from destruction. Additional game-like aspects include competition and score-points. Both the game-like and pedagogical features are incorporated as students learn 11 topics of research methodology (i.e. correlation vs. causation and generalization of results) across three separate learning modules. In these three modules, students first learn the factual information (Cadet Training module), then apply the information (Proving Ground Module), and finally generate questions (Active Duty Module). At the conference, we will present results on learning gains from interactions with this game and make recommendations for future modifications.
PRESIDENT’S SYMPOSIUM

Dr. Brendan Johns
Communicative Disorders and Sciences Department
University at Buffalo

Content-Driven Machine Learning: Using Lexical Variability to Optimize Models of Natural Language

The collection of large text sources has revolutionized the field of natural language processing, and has led to the development of many different models that are capable of extracting sophisticated semantic representations of words based on the statistical redundancies contained within natural language. However, these models are trained on text bases that are a random collection of language written by many different authors, designed to represent one’s average experience with language. This talk will focus on two main issues: 1) how variable the usage of language is across individuals, and 2) how this variability can be used to optimized models of natural language processing. It will be shown that by optimizing models of natural language based on the same lexical variability that humans experience, it is possible to attain benchmark fits to a wide variety of lexical tasks.

Evimaria Terzi
Computer Science Department
Boston University

Computational methods for team formation

The performance of a team depends not only on the abilities of its individual members, but also on how these members interact with each other. Inspired by this premise and motivated by a large number of applications in educational, industrial, and management settings, team-formation problems aim to engineer teams that are effective and successful. In the talk, we will discuss computational approaches to team-formation problems and highlight the connection of these approaches to models of social theory that capture team dynamics.
KEYNOTE

Dr. Arthur Graesser
Professor, Department of Psychology
Co-Director, Institute for Intelligent Systems
The University of Memphis

Twenty Years of AutoTutor: Computer Agents Help Students Learn by Holding Conversations in Natural Language

AutoTutor helps students learn by holding conversations in natural language. AutoTutor attempts to (a) interpret the students’ verbal contributions, actions, and in some systems their emotions and (b) generate conversation moves that advance their learning, motivation, and coverage of the subject matter. Many of AutoTutor’s conversation patterns simulate human tutoring, but other patterns implement ideal pedagogies -- that opens the door to the possibility of computer tutors eclipsing human tutors in learning gains. Current versions of AutoTutor yield learning gains on par with novice and expert human tutors. This presentation gives highlights of 20 years of research with AutoTutor and similar systems developed by my colleagues. Recent AutoTutor projects are investigating three-party conversations, called triologues, where two agents (such as a tutor and student) interact with the human student. One recent system has triologues that help struggling adult readers improve their text comprehension strategies. Two systems under development help sailors in the Navy advance their careers through training (PAL3) and acquire expertise in basic electricity and electronics (ElectronixTutor).
Notifications

Birnbaum Scholarships

If you are a first-time graduate student attending SCiP, and the primary presenter, you may be eligible to have your registration waived as part of the Birnbaum Scholarship for first-time graduate presenters. Please contact our secretary, Rick Dale (rdale@ucmerced.edu) for details.

Castellan Award

The Society sponsors The Castellan Student Paper Award for the outstanding student paper or poster. Student papers or posters on the application of computers to any area of psychology (theoretical, experimental, applied) are welcome. Eligibility is open to work done by a student currently enrolled in undergraduate or graduate courses, or work done as part of a course, thesis, or other student research by a person who graduated within the last year. The student must be the primary author and the presenter of the paper.

The winning author will receive a one-year membership in the Society, a one-year subscription to BRM, and a $200 cash prize. Students who wish to have their papers considered for the award should indicate their desire at the time of submission and are required to submit the complete manuscript for evaluation for the competition by October 1, 2016. The award for the outstanding student paper will be presented at the conference during the business meeting.

The 2015 winner of the Castellan Award was Felix Henninger:

Felix Henninger (University of Koblenz-Landau, Max Planck Institute for Research on Collective Goods, University of Mannheim), Pascal J. Kieslich (University of Mannheim), Psynteract: A flexible, cross-plattform, open solution for interactive experiments.

Special Issue of Behavior Research Methods

Conference presenters (both spoken and poster presentations) are invited to submit articles based on presentations at the SCiP meeting to a special issue of Behavior Research Methods, the official journal of the SCiP. The September 2017 issue of Behavior Research Methods will feature the 2016 SCiP conference. Although papers submitted to the special issue go through a normal review process, the time to publication is greatly accelerated. Consequently, this is an excellent opportunity to get your important research published in an extremely timely fashion. This is one of the important perks of the attending the conference. Submissions should be prepared according to the guidelines of BRM, which can be found at http://brm.psychonomicjournals.org/, and submitted through the BRM submission portal. It is important to indicate that the submission is for the “SCiP Issue.”

The deadline for article submission is December 31, 2016.
Introducing E-Prime 3

New Features in E-Prime 3.0

- New Button sub-object for response collection
- New Choice and Slider sub-objects for surveys and recalls
- New Slide Layout Templates for quick design
- Online Experiment Library
- Auto-generate text data files with columns of interest
- Interactively run List rows for runtime or debug
- Create conditional Task Events using subroutines
- Find and replace object properties in experiment
- Start experiment from any List object
- Run an experiment in floating window for debug
- Improved audio/video playback and load time
- Improved user interface

And more!

A Multifunctional Response and Stimulus Device

pstnet.com