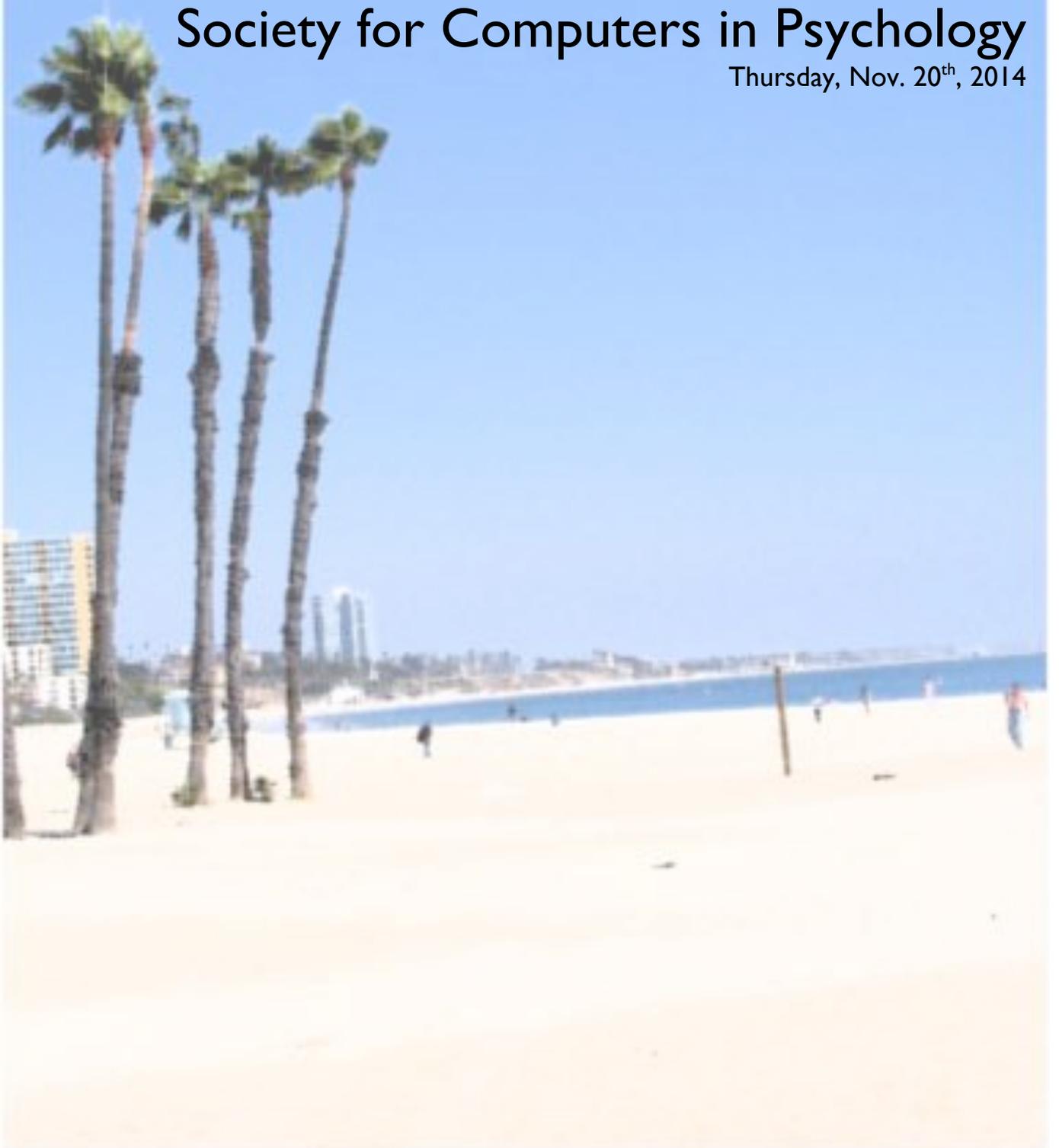


SCiP 2014
Long Beach, California
44th Annual Meeting of the
Society for Computers in Psychology
Thursday, Nov. 20th, 2014



SCiP 2014

Officers of SCiP

President: Kay Livesay
President Elect: Peter Dixon
Past President: Mike Jones
Secretary-Treasurer: Rick Dale

Steering Committee

Rick Dale, UC Merced	Monica Riordan, Chatham University
Chris Koch, George Fox University	Todd Gureckis, New York University
Xiaowei Zhao, Emmanuel College	Chris Westbury, University of Alberta
Ulf-Dietrich "Sonoma" Reips, University of Konstanz	Scott Crossley, Georgia State University
Zhiqiang Cai, University of Memphis	Jon Willits, Indiana University

Program Co-chairs

Kay Livesay & Rick Dale

Assistant Program Chair

Chris Koch

Proceedings Editors

Kay Livesay & Rick Dale

Federation Representative

Joseph Young

Notes from the Program Chair

Welcome to the 44th annual meeting of the Society for Computers in Psychology! It has been a privilege for me to oversee the formation of this year's program. I would not have been able to pull it off without Rick Dale. I am honored to have Marti Hearst as our keynote speaker, Henrike Moll and Greg Bryant as our presidential symposium speakers, and I want to extend thanks to Ulf-Dietrich Reips and Chris Wolfe for organizing our symposia this year. Finally, I want to thank the vendors for their support, please stop by their tables while at the conference. Please enjoy SCiP 2014!

Kay Livesay

Welcome from the President

Welcome to the 44th Annual Meeting of SCiP. Rick and I have worked hard to put together an interesting and varied program; I hope you enjoy the variety. I am excited about our keynote speaker, Marti Hearst. As higher education starts to evolve and colleges are forced to rethink what it means to educate a student, online education needs to be examined and evaluated for best practices. Marti Hearst is on the forefront of examining these issues.

I have been honored to be the president of SCiP for the last year. I joined as a member when I was a graduate student, became an assistant program chair in 2004, program chair in 2005 and the secretary treasurer from 2006-2009. Our society keeps evolving as the nature of computing evolves. It is this constant change that makes SCiP an interesting and exciting group of people with which to work. Our conferences always feature cutting-edge technology and new ideas about how computers can help us better understand behavior. Thank you all for joining us and please enjoy the talks and posters.

Kay Livesay

Conference Floor Map



Time	Seaview A	Seaview B	Seaview C
7:30-8:00 am	7:30 am Registration – Seaview Foyer - Refreshments (coffee and treats)		
8:00-9:15 am	<p style="text-align: center;">Statistics and Measurement</p> <p>Chair: Plant</p> <p>8:00: Love, et al. 8:15: Snow, Allen, Jacovina & McNamara 8:30: Mueller, Perelman & Veinott 8:45: Gomez, Breithaupt & Perea 9:00: Plant</p>	<p style="text-align: center;">Language Acquisition/Modeling</p> <p>Chair: Burgess</p> <p>8:00: Shaoul & Willits 8:15: Lan, Wang, Fang, Hsiao, Tzeng, & Li 8:30: Zhao 8:45: Dye & Ramscar 9:00: Burgess & Maples</p>	<p style="text-align: center;">Intelligent Tutoring</p> <p>Chair: Jacovina</p> <p>8:00: Li, Cai & Graesser 8:15: Baer, Hays, Cai & Graesser 8:30: Cai, Graesser, Hu & Nye 8:55: Allen, Roscoe, Snow & McNamara 9:00: Jacovina, Snow, Allen, Roscoe & McNamara</p>
9:15-9:30 am	Coffee Break: Foyer		
9:30-10:45 am	<p style="text-align: center;">Individual Differences/Measures</p> <p>Chair: Poddar</p> <p>9:30: Hays, Kapteyn & Liu 9:45: Tan & Mueller 10:00: Chubala & Jamieson 10:15: Glaven & Houpt 10:30: Poddar, Kattagoni & Singh</p>	<p style="text-align: center;">Semantic and Language Modeling</p> <p>Chair: Vinson</p> <p>9:30: Keith & Westbury 9:45: Szary & Dale 10:00: Gann & Matlock 10:15: Shubeck & Huetten 10:30: Vinson & Dale</p>	<p style="text-align: center;">Leveraging Online and Network Tools</p> <p>Chair: Chen</p> <p>9:30: Rodriguez, Paxton & Dale 9:45: Travers 10:00: Ciproso, Serino & Riva 10:15: Chen, Gates, Li & Proctor 10:30: <i>moved to poster session</i></p>

10:50-12:35 pm	<p style="text-align: center;">Symposium Internet-Based Research</p> <p>Organizer: Ulf-Dietrich Reips</p> <ol style="list-style-type: none"> 1. Reips 2. Dantlgraber, et al. 3. Kuhlmann, et al. 4. Stieger & Reips 5. Reips & Bannert 	<p style="text-align: center;">Measurement Developments and Applications</p> <p>Chair: Russell</p> <p>10:50: Smedema, Schulte-Mecklenbeck & Lejarrag</p> <p>11:05: Abney & Kello</p> <p>11:20: Simonsen & Fusaroli</p> <p>11:35: Huette & Viaud</p> <p>11:50: Russell, Snow, Allen & McNamara</p> <p>12:05: <i>moved to poster session</i></p>	<p style="text-align: center;">Symposium Sharable Knowledge Objects</p> <p>Organizer: Chris Wolfe</p> <ol style="list-style-type: none"> 1. Hu, Nye & Shubeck 2. Brust –Renck, et al. 3. Widmer et al. 4. Cedillos-Whynott et al. 5. Wolfe & Reyna 6. Cai 7. Graesser, discussant
12:35-1:30 pm	12:35-1:30 pm - Lunch and Poster Setup		
1:30-2:30 pm	Poster Session (Grand Ballroom, Level 2)		
2:30-3:30	<p>Presidential Session – Foundations of Social Cognition and Culture – Seaview Ballroom C</p> <p>Henrike Moll, <i>University of Southern California</i> Toddlers' Affective Anticipation of Others' Misguided Actions</p> <p>Greg Bryant, <i>University of California, Los Angeles</i> Large-Scale Cross-Cultural Experimental Psychology Research: Issues and Prospects</p>		
3:30-3:45 pm	Snack Break: Foyer – Sponsored by Metafilter		
3:45-4:55 pm	<p>Keynote Address - Seaview Ballroom C</p> <p>Marti Hearst, <i>University of California, Berkeley</i></p> <p>Towards Collaborative Learning At Scale</p>		
5-6:00-pm	Business Meeting – Seaview Ballroom C		

Presidential Session

Foundations of Social Cognition and Culture

In developing technologies and future methods for social learning, cognition and coordination, it is crucial to have a strong theoretical basis for understanding of how human cognition is organized in social contexts. The speakers for this year's symposium were selected on the basis of their groundbreaking investigation of core social cognition and culture. Prof. Moll has conducted extensive studies on the evolutionary and developmental foundations of social cognition. Prof. Bryant has conducted exciting and impressively large cross-cultural studies of human social and communicative cues, such as laughter. Together these exciting speakers will give us an important perspective on the *Foundations of Social Cognition and Culture* from multiple angles.

Introduction

Kay Livesay
Linfield College

Toddlers' Affective Anticipation of Others' Misguided Actions

Henrike Moll
University of Southern California

Large-Scale Cross-Cultural Experimental Psychology Research: Issues and Prospects

Greg Bryant
UCLA

Internet-based research: New horizons

Organizer – Ulf-Dietrich Reips

Department of Psychology, University of Konstanz, Germany

The Internet provides a powerful infrastructure for data collection and many researchers have been taking advantage of it to conduct basic and applied research. This symposium will cover new developments and present tools and examples of how to use the Internet for research. We focus specifically on 1) the rise of mobile computing (smartphones) in research, 2) tools for analyses and visualization of Internet data and effect sizes, 3) applications in eHealth and personality research.

Within the symposium, several new free Internet-based research tools (“Web apps”) for researchers will be presented. We report on a smartphone app field study about well-being that yielded over 8000 measurements within 14 days. We discuss how new on-board sensors in smartphones (e.g., GPS, gyroscope, accelerometer, thermometers) can be used for scientific research. Two of the Web apps presented in the symposium were developed in R, a free software environment for statistical computing and graphics. DropR is a Web App to analyze and interactively visualize dropout, a type of dependent variable that has become popular and challenging with the rise of Internet-based research. Log R is a package written in R suited to analyze server log files. LogR offers seamless integration with other R packages and organizes log file data into proper formats suited for behavioral and social research. The fourth app to be presented builds on recent discussions of common practices with null hypothesis testing and of “new statistics”, it turns statistical copy of effect size measurement into intuitive visual displays. All apps reported on in the symposium are provided as a free Web service from researchers for researchers.

The fifth paper reports on investigations of the German translation of the Internet Addiction Test (IAT) and an adapted version, the Facebook Addiction Test (FAT). For both tests, results favor a two-factor solution. The final talk reports on an Internet-based intervention for patients in cardiac rehabilitation and others interested in lifestyle change that carries findings on advantages of online visual analogue scales into the eHealth context. A sample of 637 participants with a mean age of 52 years was subjected to an online lifestyle intervention aimed at supporting participants over eight weeks with additional follow-up measurements. Results show that choosing the right measurement format in Internet-based interventions may lead to better health effects in the long run. Methodologies are invented or improved throughout all papers, and will hopefully provide an inspiring ground for fruitful discussion.

Factor structure of the Internet Addiction Test and Facebook Addiction Test

Dantlgraber*, Wetzel*, Schützenberger** & Stieger*

* University of Konstanz, Germany

**University of Vienna, Austria.

The Internet Addiction Test (IAT) developed by Young (1998) has been translated into several languages and is widely applied in clinical assessment. However, the factor structure of the IAT is still

unclear with some authors arguing for one factor and others for two or more factors. This paper investigates the German translation of the IAT regarding its factor structure and compares the results to the factor solutions found in other translations of the IAT as well as the original IAT. Furthermore, the factor structure of an adapted version of the IAT, which specifically assesses the addiction to Facebook (Facebook Addiction Test, FAT), is investigated. In addition, we analyze the relationship of FAT scores to frequency ratings of the engagement in several Facebook activities such as chatting, reading newsfeeds, or posting status updates. We also investigate the stability of IAT scores over a period of one week and, using latent class analysis, we examine whether qualitatively different types of Internet users and Facebook users can be differentiated.

Both the German IAT and the FAT favor a two-factor solution. Narrowing the scope of the IAT to focus on the addiction to Facebook does not appear to impair the factor structure, instead a slightly higher variance explanation can be achieved. The practical implications of our results for the clinical assessment of Internet addiction are discussed.

Using Visual Analogue Scales to increase response rates and compliance: Results from an eHealth intervention study

Kuhlmann*,**, Reips*, Wienert** & Lippke**

*University of Konstanz, Germany

**Jacobs University Bremen, Germany

Visual Analogue Scales (VAS) have been shown to be a valid measurement instrument and better alternative to Likert-type scales in online research, both empirically and theoretically (Reips & Funke, 2008; Funke & Reips, 2012). Upsides include more differentiated responses, better measurement level and less error. The present study examined VAS in the context of an online lifestyle intervention for patients in cardiac rehabilitation and others interested in lifestyle change. A sample of 637 German participants with a mean age of 52.1 years (SD = 12.5) took part in the study. The intervention aimed at supporting participants over eight weeks with additional follow-up measurements. The participants were randomly assigned to either a Likert-type or VAS version of the measures, which differed only in the answer scale formats presented. Results showed that neither means nor standard deviations of items or scales differed across the versions. However, participants in the VAS condition produced fewer missing values, on average about 10% less. Furthermore, the VAS condition exhibited significantly less dropout during the eight-week intervention period. As a consequence, the VAS may lead to better health effects in the long run and improved maintenance of lifestyle changes. These results provide further support for the use of VAS in online research, extending the scope to senior samples in the health context. As this is an ongoing intervention, we will be able to report further results during the conference.

Smartphone apps in science: How to use them in empirical research

Stieger* & Reips*

*University of Konstanz, Germany

Data collection methods in the social and behavioral sciences have always been inspired by new technologies. The introduction of the Internet had a major impact in advancing the methodological repertoire of researchers, with Internet-based experiments, online questionnaires and non-reactive online data collection methods, to name just a few. Meanwhile, the next major impact from technology is

hitting research – smartphones. The penetration rate of these small mobile devices is increasing rapidly, and they offer a multitude of new sensors that can be used for scientific research (e.g., GPS, gyroscope, accelerometer, temperature sensors). We report a smartphone app field study about well-being conducted in German-speaking countries ($N=219$). It took place for 14 days with three measurements per day (8000+ well-being judgments). Based on this study, we discuss important aspects of the planning of a smartphone study. How can a smartphone app be programmed? What are the prerequisites for a successful implementation of a smartphone study? What are the pitfalls and possible solutions? What kind of recruitment strategies are available to invite potential participants? To sum up, the presentation aims not only to present empirical data about an exemplary smartphone study, but also to present the unique aspects of smartphone studies compared to traditional research methods of data collection.

DropR: An R-based Web App to analyze and visualize dropout

Reips* & Bannert*,**

*University of Konstanz, Germany

**ETH Zurich, Switzerland

With Internet-based research non-response such as lack of responses to particular items and dropout have become interesting dependent variables, due to highly voluntary participation and large numbers of participants (Reips, 2000, 2002). In this paper we present DropR, a Web App to analyze and visualize dropout, and we develop and discuss the methodology of using and analyzing dropout in Internet-based research. The Web App was developed in R, a free software environment for statistical computing and graphics. Among other features, DropR turns input from datasets in various formats into visual displays of (1) dropout curves, (2) percent remaining, and (3) dropout trees. It calculates parameters relevant to dropout analysis, such as Chi Square values for points of difference, initial drop, and percent remaining in stable states. With automated inferential components, it identifies critical points in dropout and critical differences between dropout curves for different experimental conditions and produces related statistical copy. The visual displays are interactive, users can use mouse over and mouse drag and click to identify regions within a display for further analysis. DropR is provided as a free Web service from researchers for researchers.

Sharable Knowledge Objects: A Platform for Portable Intelligent Tutoring Systems

Organizer - Christopher R. Wolfe

Mimi University, Oxford, Ohio

Sharable Knowledge Objects (SKO, formerly AutoTutor Lite) is a platform for Web-based Intelligent Tutoring Systems (ITS). SKO uses semantic decomposition to interact with people in natural languages such as English. Tutorials created with SKO present users with images, sounds, text, and video. SKO has talking animated agents (avatars) that converse with users on the basis of expectations. A speech act classifier (SAC) analyses users' written input and compares text entered by a learner in a text box to expectation texts created by the ITS developer using latent semantic analysis (LSA) or norms from human free association tasks. It creates a semantic space from a large text corpus as a representation of the semantic relationship among words on the basis of their co-occurrences in the corpus. An ITS created with SKO compares sentences entered by a user to carefully constructed expectation texts that embody good answers. The tutor can then give appropriate feedback to students and encourage elaboration and other verbal responses on the basis of this comparison. ITS created with SKO can thus be used to encourage a variety of verbal behaviors including self-explanation and argumentation. One of the most important contribution of SKO is that it allows ITS developers to create effective tutorials including dialogues between avatars and learners without a team of experienced computer scientists as is typically needed to develop human-computer dialogues using other discourse technologies. SKO has been used successfully to create intelligent tutorials on photography, civics, medical decision-making, STEM topics, obesity prevention, and other knowledge domains. In this symposium, Xiangen Hu – the creator of SKO – and colleagues present on a system that holds conversations with learners in natural language while they learn Algebra I. Priscila Brust-Renck describes EatFit, a system that promotes nutrition and exercise to prevent obesity. Colin Widmer presents data on tutorial dialogues promoting "gist explanations" of genetic testing and breast cancer risk. Elizabeth Cedillos-Whynott describes efforts to encourage users to develop pro and con arguments about genetic testing for breast cancer risk in an ITS and their consequences for learning and decision-making. Wolfe and Reyna draw on Fuzzy Trace Theory to discuss the theory and practice of promoting "gist explanations" and present data on the effectiveness of this approach. Finally, Zhiqiang Cai presents a framework for integrating conversations into Intelligent Tutoring Systems. Renowned ITS expert Art Graesser will serve as the symposium discussant.

Shareable Knowledge Objects as Portable Intelligent Tutoring Systems Modules: Integrating Computerized Natural Language with STEM Topics.

Hu^{^*}, Nye^{*} & Shubeck^{*}

[^]Central China Normal University

^{*}The University of Memphis

Sharable Knowledge Objects (SKO's) are portable Intelligent Tutoring Systems (ITS) modules that help teachers teach and students learn Science, Technology, Engineering, and Mathematics (STEM)

topics. We are developing and testing these SKO's as one of the four STEM Grand Challenge projects funded by the Office of Naval Research. The project develops an enhanced version of AutoTutor that uses animated conversational agents to hold conversations with learners in natural language while they learn the knowledge, procedures, and skills of Algebra I. The goal is to produce efficient and cost-effective ITS modules on STEM topics that are comparable with experienced human tutors. The presentation will cover the following aspects of this project: (1) relevant **theory of learning**, such as the expectation-misconception tailored dialog in tutoring, (2) signature **enabling technologies** such as semantic representation analysis, (3) useful **implementation details** such as assembling mathematics content and implementing cognitive task analysis methods, and (4) empirical **efficacy studies** to test advanced learning technology in classrooms.

Testing the Efficacy of an Intelligent Tutoring System Grounded in Fuzzy-Trace Theory to Prevent Obesity

Brust-Renck*, Reyna*, Wolfe** & Morant*

*Cornell University

**Miami University

We used Sharable Knowledge Objects (SKO) to create an Intelligent Tutoring System (ITS) grounded in Fuzzy-Trace Theory to teach women about obesity prevention. FTT predicts that reliance on gist representations (as opposed to verbatim) is more effective in reducing health risks; we extended these predictions to nutrition and fitness. Technical information was translated into decision-relevant gist representations and gist principles. SKO also engage women in dialogue and "understands" them using Latent Semantic Analysis. The effectiveness of the ITS was tested in a randomized, controlled experiment. Participants were randomly assigned to either the obesity prevention tutorial or a control tutorial on a different content using the same technology. Participants were then given a test of content knowledge about nutrition and fitness, a gist comprehension measure, and questions about their attitudes, intentions of healthy behavior, and gist principles. The intervention group performed significantly better than the control group on content knowledge about nutrition and fitness, gist comprehension, attitudes, intentions of healthy behavior, and endorsement of healthy gist principles. We also analyzed SKO interactions using reliable scoring rubrics for content and response quality. An analysis of the tutorial dialogues revealed significant correlations between the quality of the conversations and the scores on subsequent knowledge tests, endorsement of healthy attitudes, intentions of healthy behavior, and gist principles. The new theory-based ITS may be fruitfully applied to assist laypeople in making preventive health decisions by effectively teaching content knowledge, enhancing endorsement of healthy gist principles, and increasing the intention to perform healthy behavior.

Tutorial Dialogues and Gist Explanations of Genetic Breast Cancer Risk

Widmer**, Wolfe**, Reyna*, Cedillos-Whynott** & Weil**

*Cornell University

**Miami University

Much work has been done to show the impressive learning outcomes of intelligent tutoring systems, particularly those that are able to interact with learners in natural language dialogues. One such tutor is

the Breast Cancer Genetics Intelligent Semantics Tutor (BRCA Gist), a tutor developed using the web-based SKO that is designed to teach women about genetic risk of breast cancer. Several studies have shown the efficacy of BRCA Gist as a whole with multiple populations. Three experiments also provide rich data illustrating how natural language dialogues impact learning outcomes. Dialogue scripts of participant verbal interactions with BRCA Gist from three studies were rated for the content covered in the dialogue by participants and for the appropriateness of the tutor's responses using a rubric. The human-rated rubric scores for content covered by participants were strongly correlated with BRCA Gist's own internal coverage scores for the dialogues, indicating that the tutor is appropriately able to assess dialogues for content. The tutor responded appropriately to participants more than 90% of the time. Participant performance during the tutor dialogues, as assessed by both the tutor and the human ratings, was consistently associated with learning outcomes about genetic risk of breast cancer. Dialogue content coverage was highly predictive of performance on a multiple choice declarative knowledge test, as well as moderately predictive of participant's performance on a task measuring gist understanding of genetic risk of breast cancer. Dialogue content coverage did not consistently predict ability to correctly classify women on genetic risk.

The Case Against Genetic Testing: The Effectiveness of Argumentation in an Intelligent Tutoring System

Cedillos-Whynott**, Wolfe**, Reyna*, Widmer** & Weil**

*Cornell University

**Miami University

The effectiveness having learners develop arguments aided by an Intelligent Tutoring Systems (ITS) has received little research attention. We examined the use of argumentation in the BRCA GIST ITS helping people make arguments for and against genetic testing for breast cancer risk. In three studies, 190 women recruited on-line, from two universities, and a community in upstate New York participated. Using Toulmin's model of argumentation, two independent raters judged participants' verbal responses when asked to make a case for and against genetic testing for breast cancer risk (IRR = .85). Despite prompting from an avatar, when asked to make a case for genetic testing interactions, 53% did not make an argument, while only 47% presented minimal argument elements implicitly or explicitly. In the case against genetic testing interaction, 56% did not make an argument, while only 44% met minimal criteria for an argument. However, those who made arguments against (but not for) genetic testing performed significantly higher on subsequent learning outcome measures. Across studies some differences were found between strength of arguments when making a case against genetic testing and measures of gist comprehension and the ability to categorize genetic risk. Additionally, when making a case against genetic testing, those that made arguments performed better on content knowledge, risk categorization, and gist comprehension. We found significant correlations between ratings of the quality of arguments against genetic testing and coverage of content, knowledge test scores, risk categorization, and gist comprehension. We discuss implications regarding the potential effectiveness of argumentation in ITS.

Promising Practices for Encouraging Gist Explanations in Tutorial Dialogues

Wolfe** & Reyna *

*Cornell University

**Miami University

The value of self-explanation in academic learning has been well documented. Both expert human tutors and Intelligent Tutoring Systems (ITS) such as AutoTutor encourage learners to construct richer, more elaborate, and more causally connected explanations. However, less is known about the potential for ITS to improve decision making, and best practices for creating ITS-based decision aids in domains such as medical decision making. Drawing upon Fuzzy-Trace Theory (FTT), we propose a set of promising practices for encouraging gist explanation in ITS to improve decision making. Supported by experimental research, FTT places a premium on helping people develop useful and appropriate gist representations. This suggests a number of promising practices to implement when ITS's are used as decision aids in helping people form good gist explanations. As with other theories, FTT indicates a key role for self-explanation. However, FTT suggests that people should be encouraged to develop explanations in their own words, rather than trying to precisely reproduce to-be-understood information verbatim. Second, FTT suggests that the focus of prompts and hints should be decision-relevant dimensions of the task, rather than more open-ended prompts or prompts for specific detailed information. Finally, some dialogues should deliberately focus on the consequences of decision alternatives. An ITS called BRCA Gist, the first ITS applied to patients' medical decision making, employs these emerging practices to help women understand and decide about genetic testing for breast cancer risk. We provide evidence for the effectiveness of gist explanations promoted by these practices and discuss future directions.

AutoTutor Framework for Integrating Conversations into Intelligent Tutoring Systems

Cai, Graesser, Hu, & Nye

Institute for Intelligent Systems, University of Memphis

AutoTutor is an established framework that supports conversions for intelligent systems, with an emphasis on intelligent tutoring systems. The AutoTutor framework is based on extensive research on tutoring, natural language processing and discourse analysis. Recent advances enable AutoTutor to more widely integrate with other systems. The major components of AutoTutor framework include ACE (AutoTutor Conversation Engine), ASAT (AutoTutor Script Authoring Tool), and a set of AutoTutor delivery interfaces. ACE interprets AutoTutor scripts containing conversation transition rules and conversation content. ACE is responsible for assessing the quality of user responses, interpreting conversation transition rules, and sending the client program action sequences to execute. ACE also provides detailed interaction logs which can be used by the client to make immediate tutoring decisions or to record for researchers studying the conversation sessions later. Conversational scripts for ACE are designed using ASAT, which has several different forms. ASAT-D is a desktop version that is most powerful and flexible, supporting the design of challenging dialog rules. This version is primarily used by specially-trained authors. ASAT-W is a web version. While it provides limited conversational rule authoring, it facilitates online collaborative authoring of simpler conversation scripts. ASAT-V is a visual version of ASAT where conversational rules are designed using flow charts, which greatly reduces the difficulty of conversation rule design. AutoTutor client interfaces exist as a component in

ASAT and also as standalone web applications. Through this set of tools, AutoTutor framework makes it possible to integrate intelligent conversations into a variety of systems.

Discussion.

Art Graesser. University of Memphis



Keynote Address

Marti A. Hearst
University of California, Berkeley

Towards Collaborative Learning At Scale

Massive Open Online Courses (MOOCs) have emerged as a new and provocative approach to education. MOOCs as initially deployed incorporate an innovative learning technique, at least compared to the standard computer science in-person classroom, consisting of sequences of very short lecture segments followed by student learning activities. The education literature shows strong support for the pedagogical value of pausing to check student comprehension after short presentations of material, as the MOOC active learning exercises do. The literature also shows, though, that the effects are even stronger if students engage with the material in small groups before progressing to the next set of material. This practice is variously known as peer learning, collaborative learning, and cooperative learning, and has been studied both in the classroom and online in the field of Computer Supported Collaborative Learning, but it has never been studied at scale. At the same time, MOOCs are not set up for real-time collaboration. In our research, we are putting the two ideas together, and are working towards the goal of improving retention and learning in MOOCs. This talk will discuss early-stage research into the incorporation of collaborative learning at scale, with the goal of incorporating it into MOOCs.

Joint work with Bjoern Hartmann, Armando Fox, and Derrick Coetzee

Bio: Dr. Marti Hearst is a professor in the School of Information at UC Berkeley, with an affiliate appointment in the Computer Science Division. Her primary research interests are user interfaces for search engines, information visualization, natural language processing and, more recently, improving MOOCs. She was recently named a Fellow of the ACM, and has won three departmental Excellence in Teaching Awards. She is also known for the book *Search User Interfaces* and for the Flamenco project which advanced faceted navigation as a standard search technique, for lexico-syntactic patterns for ontology discovery (“Hearst patterns”), and the TextTiling discourse segmentation technique.

- 8:00** JASP; A free and open alternative to SPSS, a rich and user friendly alternative to R.

Love, Selker, Smira, Verhagen, Steingrover, Morey & Wagenmakers

The two dominant software platforms for performing statistical analyses in psychology are SPSS and the R programming language. SPSS is expensive, closed source, and does not implement the newest analyses (such as recent developments in Bayesian statistics). In contrast, the R programming environment is open source and implements the newest analyses, however it requires users to write software in a programming language, and is a challenging platform for people without programming experience to use. This talk introduces JASP, an alternative to both SPSS and R which combines the best of both worlds. JASP implements the latest analyses, is released under an open source license and provides a rich user interface such that individuals without programming experience can use.

- 8:15** Does agency matter?: Path analysis within a game-based system

Snow, Allen, Jacovina & McNamara

When students exhibit control and employ a strategic plan of action over a situation they are said to be demonstrating *agency* (Bandura, 2001). The current work is comprised of two studies designed to investigate how agency manifests within students' choice patterns and ultimately influences self-explanation quality within the game-based system iSTART-2. In Study 1, 75 college students interacted freely within iSTART-2 for two hours. Using random walks and Entropy analyses, we visualized and quantified students' choice patterns as either controlled or random and examined how this related to self-explanation performance within iSTART-2. Findings revealed that students who demonstrated controlled choice patterns generated higher-quality self-explanations compared to students who exhibited random choice patterns. We hypothesized that the advantage for students who exhibited a controlled pattern was in part driven by their experience of agency. That is, engaging in controlled patterns should be advantageous only when doing so is a result of students' strategic planning. In Study 2, we tested this hypothesis by assigning 69 students a choice pattern (i.e., controlled or random) that had been yoked to students from Study 1, thus removing students' ability to exert agency over their learning. Findings from this study revealed no differences in self-explanation quality between the groups assigned to random and controlled choice patterns. Collectively, findings from these studies support the notion that success within game-based systems is related to students' ability to exert agency over their learning paths.

- 8:30** An optimization approach for measuring the divergence and correspondence between paths.

Mueller*, Perelman* & Veinott**

* Dept. of Cognitive and Learning Sciences, Michigan Technological University

** Applied Research Associates, Inc.

Spatial paths are produced or observed in many areas of scientific research, and are of interest in a number of behavioral data analysis, machine intelligence, machine vision, and data mining applications. Within psychological domains, such data are produced in a number of paradigms, including map path communication (Anderson's map task), spatial search and pathfinding (Mueller et al., 2013), traveling salesman solution (Pizlo et al., 2006), eye movements, mouse-tracking, body trajectory monitoring, and other areas. We describe a new proposed Algorithm for finding the Least-Cost Areal Mapping between Paths (ALCAMP), which robustly measures the area-based distance between two paths and also produces a mapping between points on the two paths. The method works by optimizing a least-cost mapping (in terms of areal deviation) between two paths, and is robust to a number of aspects in real path data, such as crossovers, self-intersections, differences in path segmentation, and partial or incomplete paths. Unlike similar algorithms that produce distance metrics between trajectories (i.e., paths that include timing information), this algorithm uses only the order of observed path segments to determine the mapping. We describe the algorithm and show its results on a number of sample problems and data sets in psychological and behavioral domains. We also describe available software code written in the R statistical computing language that implements the algorithm.

Corresponding author contact: shanem@mtu.edu

8:45 On the application of survival analyses to the timing of cognitive processes.

Gomez*, Breithaupt** & Perea**

* DePaul University

** Universitat de Valencia

The uncovering of the time course of the influence of different factors in human performance is one of the principal topics of research in cognitive psychology/neuroscience. Over the past decades, researchers have proposed several methods to tackle this question using latency data. This presentation focuses on a recent procedure proposed by Sheridan and cols. (Reingold, Reichle, Glaholt, & Sheridan, 2012, Sheridan et al., 2012). They employed hazard functions and survival analyses on fixation durations to provide "precise estimates" of the timing of the first discernible influence of variables (e.g., word frequency). Because the method is intriguing and promising, an exploration of its strengths, biases and shortcomings is in order. Here we report the results of systematic simulations directed to parameter recovery that revealed that this method tends to over-estimate the divergence point with a realistic number of observations per condition.

Corresponding author contact: pgomez1@depaul.edu

9:00 Quick, quick, slow: Timing inaccuracy in computer-based studies means we may need to make use of external chronometry to guarantee our ability to replicate.

Plant

The Black Box ToolKit

Many researchers make use of computers to present stimuli, interface with other equipment and record responses. There is a tacit understanding that achieving consistency may not be as easy as simply entering identical parameters into different hardware configurations or software packages.

With renewed emphasis on replication across the field now might be an appropriate juncture to address the issues at the individual researcher, journal editor and funder levels. Our own research suggests accuracy is getting worse as hardware becomes faster and software for experimental control more complex. In most other fields poorly controlled studies that did not outline methodology fully or state confidence limits for equipment accuracy would not be published.

Our new methodology and accompanying hardware means that researchers can be sure of presentation, synchronization and response timings because each is measured using external equipment, sensors and response devices which is separate to the computer, tablet or phone administering the experiment. Our new integrated response device functions as a HID keyboard meaning that it can be used with any experiment generator with all timing data accessed via Bluetooth or through a flash drive letter. For synchronization and event marking various TTL interfaces to other equipment are offered.

In this paper we outline the issues, common pitfalls and offer practical solutions, e.g. to move all presentation, synchronization and response timing off the experimental platform and onto external chronometry whilst adding little or no overhead for the experimenter. For the sake of replication we feel this is the only way forward.

Corresponding author contact: r.plant@blackboxtoolkit.com

Individual Differences/Measurement

9:30-10:45 am

Seaview Ballroom A

9:30 Use of online panels to conduct surveys

Hays*, Kapteyn* & Liu**

* University of California, Los Angeles

** University of Southern California

Use of online panels to collect survey data is increasing because it is exceedingly efficient in terms of cost, ability to recruit diverse samples quickly, and time to get data back for analysis. Most panel vendors rely on non-probability based recruitment and members of these panels are known to differ from non-members (e.g., panel members tend to be more educated and have higher socioeconomic status). Panel samples typically have unknown representativeness of the underlying population. Hence, many users of online panels utilize a quota sampling approach by targeting respondents with particular characteristics and using post-stratification adjustments (weights) to compensate for non-response and non-coverage. A comparison of responses to the PROMIS global health items across four surveys found comparable estimates of physical and mental health despite differences in survey sampling (probability vs. non-probability) (Riley et al., 2014). In contrast, Yeager et al. (2011) concluded that probability sample surveys were consistently more accurate than non-probability sample surveys, even after post-stratification weighting of the data. A study that recruited U.S. adults from 7 panel vendors using identical quotas found variability in response rates and estimated that different panel vendors appeared to draw 15-25% of their samples from a common pool (Craig et al., 2013). This paper provides an

overview of the potential benefits and downsides to use of internet panels for conducting survey research.

Corresponding author contact: drhays@ucla.edu

9:45 Adapting cultural mixture modeling to identify cultural groups based on response time patterns.

Tan & Mueller

Michigan Technological University, Department of Cognitive and Learning Sciences

Previous research (e.g., Cultural Consensus Theory: Romney, Weller, & Batchelder, 1986; Cultural Mixture Modeling (CMM): Mueller & Veinott, 2008) has used overt response patterns (i.e., questionnaires and surveys) to identify whether a group shares a single coherent attitude or belief set. Yet many domains in social science have focused on implicit attitudes that are not apparent in overt responses but can still be detected via response time patterns. We sought to adapt CMM to identify shared knowledge measured through the timing of responses that all respondent agree on, rather than actual responses themselves. We will describe the application of these new models, and report the results of simulation and behavioral studies (a geographic pair learning task) that show the sensitivity of the method to known group structures. Results reveal that the ability to recover and identify shared belief groups depends on 1. the level of noise in the measurement, 2. the differential signal for strong versus weak attitudes, and 3. the similarity between group attitudes. Consequently, the method shows promise for identifying latent groups amongst a population whose overt attitudes do not differ, but whose implicit or covert attitudes or knowledge may differ.

Corresponding author contact: yinyint@mtu.edu

10:00 An automated technique for controlling Type I error inflation in individual subjects analysis

Chubala & Jamieson

University of Manitoba

Lorch and Myers (1990) demonstrated that traditional analyses of repeated measured designs make inappropriate use of error terms, resulting in inflated Type I error rates. They proposed an alternative, two-stage approach. First, each participant's score on some dependent variable, y , is regressed on multiple predictors, $x_1 \dots x_n$, to yield individual slope estimates. Second, a hypothesis test is conducted on the coefficients for each predictor, collapsing across participants. Thus, the method not only speaks to the reliability of predictors at the group level, while controlling Type I error rates, but also yields information about individual participants for free. This represents a powerful tool for researchers seeking to untangle sources of variability in their datasets, particularly where individual differences may be indicative of separable behavioural strategies rather than noise. Despite its demonstrated empirical usefulness (e.g., Johnstone & Shanks, 1999), the method is underused in experimental psychology, we think for practical reasons. Namely, conducting regression analyses for each of N participants is time-consuming, tedious, and prone to error. We address the issue by automating the technique, providing code in both the SAS and R computing languages. The program automatically conducts N multiple regression analyses, performs hypothesis tests on each predictor, and extracts the pertinent information into easy-to-read output. We provide examples of experimental scenarios that may benefit from this powerful technique, with the hope that solving

the implementation issues will encourage broader use.

Corresponding author contact: umchubal@myumanitoba.ca

10:15 Using adaptive psychophysical methods to individualize studies involving Systems Factorial Technology

Glavan & Houpt
Wright State University

Systems Factorial Technology (SFT; Townsend & Nozawa, 1995) provides procedures for determining the architecture (e.g., parallel, serial, coactive) and stopping rule of a cognitive process that incorporates multiple sources of information. In order to reliably measure the function used to infer architecture, the researcher must selectively speed up and slow down the processing of each source of information by manipulating the saliency of the sources. While the same saliency manipulations may be suitable for all subjects in some experimental designs, individual differences are often great enough to necessitate the estimation of individualized signal intensities corresponding to low saliency and high saliency for each subject before they participate in the SFT portion of the experiment. Because the time a researcher has to collect data from a subject is limited, an adaptive psychophysical method is essential to minimizing the time spent parameterizing the experiment to the individual subject. The Psi method (Kontsevich & Tyler, 1999) has been identified as suitable for pairing with SFT because it efficiently estimates the location and slope of a psychometric function, both of which may vary across individuals. We describe an implementation of the Psi method in Python developed specifically for integration in the open source package PsychoPy (Peirce, 2007). We further discuss its use in an SFT study that explored how color and shape are used together in visual search.

Corresponding author contact: glavan.3@wright.edu

10:30 Personality mining from social media with the "Adjectival-Marker" Technique

Poddar, Kattagoni, & Singh
International Institute of Information Technology, Hyderabad

The last decade has witnessed significant work in personality mining from lexical cues in social media data. Most of this work involves a large crowd of researchers leveraging LIWC-based approaches (which primarily focuses on function words). By the means of this paper we intend to introduce an entirely unique method of personality mining from social media data called "Adjectival-marker Technique". This method draws heavily from the last 78 years of work in lexical psychology and the Big5 personality test. The method is not only a computational variant of the primordial theories of lexical psychology, but also challenges the other recent personality prediction models by conferring accuracy tending to 100% in majority of the Big5 traits.

The theories of psychology were influenced by various revolutionary concepts. For instance, "trait" - a theoretical construct which describes a basic dimension of a person's personality (Allport, 1937). These sprouted out of the Lexical Hypothesis of psychology (worked upon by Klages: 1926/1932, Cattell: 1943b, Norman: 1963, and Goldberg: 1982) which states: *"Those individual differences that are most salient and socially relevant in peoples' lives will eventually become encoded into their language; the more important such a difference, the more likely is it*

to become expressed as a single word". As a result of the computational accuracies unraveled by the above described model we also propose an addition to this Lexical Hypothesis, which could be a huge advance in determining effective personality mining from social media in the near future.

Corresponding author contact: shivani.poddar92@gmail.com

8:00 Modeling multiple aspects of language acquisition using Naïve Discrimination Learning

Shaoul* & Willits**

* University of Tuebingen

** University of Indiana, Bloomington

Language acquisition is a complex process, involving a number of different interacting skills and types of knowledge. These include learning about a language's sound system, learning the meanings of those sounds, and learning how to interpret complex and meaningful linguistic utterances. A number of different computational models have been proposed for explaining individual components of the language acquisition process, but few models or theoretical approaches are able to use the same principles to explain the multiple sub-skills that are necessary for learning a language. We will demonstrate a Naïve Discrimination Learning (NDL) model that is able to explain language acquisition simultaneously at multiple levels. Naïve Discrimination Learning is a biologically plausible, online processing model of learning. In contrast to traditional models of language acquisition, which attempt to explain how complex linguistic utterances are composed by combining smaller elements, Naïve Discrimination Learning works in the reverse, by learning to use the information in a stream of linguistic events to discriminate between the possible messages that can be encoded in these events. We will show that an NDL model, when trained on naturalistic child-directed speech from the CHILDES corpus, can be used to investigate the learning of both sub-lexical and lexical patterns in language. The ease by which our model can be used to explain a number of different language acquisition processes at multiple levels of linguistic analysis suggests that NDL has the potential to be a powerful unified model of language acquisition.

Corresponding author contact: cyrus.shaoul@uni-tuebingen.de**8:15** L2 word learning through virtual environment versus audio-visual paired association

Lan, Wang, Fang, Hsiao, Tzeng, & Li

A recent interest in language acquisition has been on how learning contexts may impact the success of a second language (L2). Traditional classroom setting involves the student's repetition of rules and memorization of words in a memory-intensive fashion, often out of touch with the natural environment in which language is actually used by native speakers (L1 speakers). However, learning L2 in a natural environment is not always available (e.g., which requires students to study abroad in the country in which the language is spoken). Recent evidence of interactive virtual environment in digital platforms may provide simulated rich contexts comparable to a natural learning environment.

Using a longitudinal training paradigm, our study examined two groups of English monolinguals who either learned Chinese words as L2 in a virtual environment in which the learner explores the language learning context (the Exploration or EXP group) or in a association-based computer presentation format in which the learner hears a Chinese word and sees a picture of the object to which the word refers (the Association or ASC group). The EXP group used Second Life for

learning (www.secondlife.com), an online virtual environment platform with 3D objects and interactive avatars. Here we report findings from both behavioral and brain measures of participants before and after training, showing that virtual environment may provide a more effective learning platform as compared with traditional associative paired learning. Our results have significant implications for second language learning, teaching, and education.

Corresponding author contact: Not provided

8:30 Investigating “Noun bias” based on CLEX database.

Zhao
Emmanuel College

The present study focuses on examining the so-called “Noun-bias”, a hypothesized universal pattern that shows predominance in the number of nouns as compared with other categories of words, in the child’s early lexical composition. Previous research has shown strong support for Noun-bias in the language development of English (Gentner, 1982), but debates exist on if this pattern can be found in some other languages, such as Chinese (Tardif, 1996). In the present study, I examine the lexical compositions of a couple of categories in children’s early vocabularies across multiple languages (N=10). Particularly, the noun-verb ratios are calculated and compared across different age groups in the ten languages. The analyses are based on CLEX (Cross Linguistic Lexical Norms, <http://www.cdi-clex.org/>) – an online database providing a collection of children’s early lexical development norms, which were originally collected by researchers using various language versions of the MacArthur-Bates Communicative Development Inventories (CDIs). The results are further discussed along with our previous analyses based on children’s speech (Liu, 2007; Mauer & Zhao, 2013) and computational simulations (Zhao et al, 2008). The purpose is to provide a complete picture of the important issue in child language development.

Corresponding author contact: zhaox@emmanuel.edu

8:45 Refining the distributional hypothesis: A role for time and context in semantic representation

Dye* Ramscar**
*University of Indiana
** University of Tübingen

Distributional models of semantics assume that the meaning of a given word is a function of the contexts in which it occurs. In line with this, prior research suggests that a word’s semantic representation can be manipulated – pushed toward a target meaning, for example – by situating that word in distributional contexts frequented by the target. Left open to question is the role that order plays in the distributional construction of meaning. Learning occurs in time, and it can produce asymmetric outcomes depending on the order in which information is presented. Learning theory predicts that systematically manipulating a word’s preceding context should more strongly influence its meaning than should varying what follows. We find support for this hypothesis in two experiments in which we manipulated subjects’ contextual experience with both high and low frequency words, while varying the location of manipulation. We consider the implications for various modeling approaches.

Corresponding author contact: meldye@indiana.edu

9:00 The dissociation of semantic factors from other lexical factors in the HAL memory model.

Burgess & Maples
University of California, Riverside

There is considerable evidence that high-dimensional memory models, such as HAL or LSA, bootstrap statistical information from the language input and that the representations formed from this process account for a wide variety of semantic and associative memory effects (Burgess & Lund, 2000; Lund & Burgess, 1996). The HAL model has also accounted for meaning relationships related to thematic, syntactic, and grammatical variables (Audet, Burgess, & Driessen, 1999) and vocabulary skill (Devitto & Burgess, 2004). More recent work has investigated meaning in dolphin communication (McCowan, et al., 2008) and syntax and structure in humpback whale song (Kaufman, Green, Seitz, & Burgess, 2012). This range of effects by a lexical co-occurrence model raises the question: Are these effects truly part of meaning representation or are they more general semantic/lexical effects that would also be associated with sub-semantic variables such as phonology, orthography and morphology. In this study we used the data from the English Lexicon Project (Balota, et al., 2007) to investigate the relationship of two semantic variables from the HAL model (semantic neighborhood density and semantic vector density) to non-semantic factors. The key findings were that neighborhood and vector density were strongly correlated with lexical decision times (.41, -.38) and less correlated with word naming times (.14, -.18). This finding is consistent with the word recognition literature that has shown that the lexical decision task is more sensitive to semantic level effects than the naming task. Additionally, the two semantic variables were not correlated with orthographic and phonological variables.

Corresponding author contact: curt@ucr.edu

Semantic Modeling

9:30-10:45 am

Seaview Ballroom B

9:30 Performance impact of full morphological decomposition and stop-lists on corpus-based semantic space models

Keith & Westbury
University of Alberta

Corpus-based semantic space models, which primarily rely upon lexical co-occurrence statistics, have proven effective in modeling and predicting human behaviour in a number of experimental paradigms that explore semantic memory representation. However, the most widely studied extant models are strongly influenced by orthographic word frequency. This has the implication that closed class words, or function words with very high orthographic frequency, can potentially bias co-occurrence statistics. As these closed class words are purported to carry primarily syntactic, rather than semantic, information, performance of corpus-based semantic space models may be improved by excluding closed class words (using stop-lists) from co-occurrence statistics, while retaining their syntactic information through other means (e.g. part of speech

tagging and/or affixes from inflected word forms). Additionally, very little work has been done to explore the effect of employing morphological decomposition on inflected forms of words in corpora prior to compiling co-occurrence statistics, despite (controversial) evidence that humans perform early morphological decomposition in semantic processing. This study explored the impact of these two factors on corpus-based semantic space models. *Results.* Based on this study, morphological decomposition appears to significantly improve performance in word-word co-occurrence semantic space models. An overall decrease in performance was observed in models employing stop-lists (e.g. excluding closed class words). Further, some evidence was found that weakens the claim that closed class words supply primarily syntactic information in word-word co-occurrence semantic space models.

Corresponding author contact: jkeith1@ualberta.ca

9:45 Modeling the Effects of Interaction and Resource Landscape on Search

Szary & Dale
University of California, Merced

Remembering can be conceptualized as a search process across the mental landscape. The success of this search process is modulated by the interaction of factors including social context and resource distribution, where resources are the target items of remembering. Rhodes and Turvey (2007) showed that the dynamics of memory search are analogous to animal foraging behaviors. Specifically, they found evidence for Lévy processes, which are similar to random walks but with path lengths that fit a specific, heavy-tailed distribution. Lévy flights are thought to optimize biological foraging under certain conditions (Viswanathan, 1999). Participants in a category recall task performed better when the distribution of their inter-response intervals was more Lévy-like (Rhodes & Turvey, 2007). Our own work adds to this finding with manipulations of social context and resource distribution. Our results suggest that groups outperform individuals in some categories, but not others, and that across all conditions performance is associated with Lévy characteristics. In order to tease apart the effects of these different parameters, we constructed a computational model to simulate foraging. In our model, two agents move across a space searching for resources. The first parameter of interest is the strength of interaction between the agents. The second set of parameters includes sparsity and clustering, which together change the patterns of resource distribution in the environment. This model allows us to explore the effects of social and environmental context on both the processes (Lévy dynamics) and products (number of resources found).

Corresponding author contact: jszary@ucmerced.edu

10:00 Machine classification of climate messages in partisan media

Gann & Matlock
University of California at Merced

Climate change is one of the most important scientific, social, and political problems of our time. Confronting the human dimension of this challenge requires a much deeper understanding of how people from all sectors communicate and make sense of issues around climate change. Some research has analyzed discourse about climate change, including information in newspaper articles, but rarely is detailed attention given to linguistic content, especially across

political domains. To gain a better understanding of how people react to climate messages in the popular media, we need sophisticated machine learning algorithms that are capable of coding and analyzing specific linguistic content of climate messages. In the current project, we expand our previous work (Gann & Matlock, 2014; under review) by using high dimensional semantic analyses to study climate discourse in conservative and progressive news media. In doing so, we explore various methods for doing automated comparisons across a large-scale diverse corpus (e.g., *National Review*, *The Nation*) and focus on particular “low-level” linguistic forms. One goal of this project is to further develop our understanding of the contrasting grammatical and semantic elements that characterize the differences between opposing partisan texts. Another is to demonstrate how our large-scale set of corpora can be used as training data for classification algorithms with the capacity to code novel data, for instance, to code written reactions to articles as either progressive or conservative.

Corresponding author contact: tgann@ucmerced.edu

10:15 Predicting Meme Success with Linguistic Features in a Two-Layer Backpropagation Network

Shubeck & Huette
University of Memphis

The challenge of predicting meme success has recently gained attention from researchers, largely due to the increased availability of social media data. Many models focus on structural features of online social networks as predictors of meme success (Weng, Menczer, & Ahn, 2014; Hogg & Lerman, 2012). Many of these analyses use the structure of the social system to explain meme behavior, but neglect the role of the meme’s impact at the local level of individual users. Other models focus on user influence within the network as a predictor of meme success (Romero, Meeder, & Kleinberg, 2011). The current work draws on theory from connectionist networks to learn, at the local level of an individual user, which linguistic features impact global meme success. We propose a new approach that involves grounding variables of interest in theories of working memory limitations, emotions, memory, and psycholinguistics. A collection of 268 memes were pulled from a meme categorizing website and were used as inputs in a multilayer backpropagation neural network. The linguistic content of each meme was analyzed with linguistic analysis tools, LIWC and Coh-Metrix (Pennebaker, Francis, & Booth, 2001; Graesser, McNamara, Louwerse, & Cai, 2004). The memes were coded as 14 string binary vectors and were then used to train the network’s connections. Once the network converged, a set of 16 new memes was used to test the generalization and accuracy of the network. Our preliminary results indicate the network is generalizing the linguistic features in order to predict success at greater than chance levels (62.5% accuracy).

Corresponding author contact: kshubeck@memphis.edu

10:30 Information and social network analysis in a large natural corpus

Vinson & Dale
University of California, Merced

Information theoretic measures have been used to analyze sequences of words across a variety of linguistic corpora. In most cases the information contained in a word is defined as a function of its local context (e.g., the probability of that word's occurrence conditioned on the preceding

word). One recent theory suggests users structure their utterances so as to maintain a uniform distribution of information across a message (e.g., a Uniform Information Density). Interestingly, a recent analysis of over one hundred thousand Word of Mouth customer business reviews focusing on user related variables, such as the average overall business review rating, revealed the information and uniformity of information within a message varies dependent on a user's intended message valence. Using this same dataset, provided its massive size and structure, the current work focused on how a user's social network structure affects the amount and uniformity of information across a user's reviews. Applying network analysis methods— now possible provided by the massive size of the dataset— the current work explores what aspects of a user's social network (e.g., clustering coefficient, transitivity, sparsity, etc.) impact a review's informational context. We discuss how methods from information theory and network theory can be leveraged to explore and explain the relevance of larger datasets in understanding certain cognitive processes.

Corresponding author contact: dvinson@ucmerced.edu

Measurement Developments and Applications

10:50-12:35 pm

Seaview Ballroom B

10:50 Eye-tracking for the masses

Smedema*, Schulte-Mecklenbeck**, & Lejarraga**

*Indiana University

** Max Planck Institute for Human Development

Eye tracking is one of psychology's favorite tools to study processing of information and attention. Eye-tracking experiments are noninvasive, easy to set up, and useful for a wide array of applications. One downside, however, is the large price tag that comes with proprietary systems, often exceeding tens of thousands of dollars. The recent availability of inexpensive eye-tracking units offered by The Eye Tribe (www.theeyetribe.com) has drastically reduced the cost of individual trackers, promising a new area of research: simultaneously tracking many participants. Although this type of research is no longer prohibitively costly, some technological barriers remain. Although Python is rapidly becoming the language of choice for scientific computing, there is currently no Python client for The Eye Tribe. We present a client that serves as a full implementation of the Eye Tribe's API as well as offering a few higher-level functions, such as a full calibration. It is lightweight, with no dependencies outside of Python's built-in functions, and is compatible with both Python 2.7 and Python 3. Further, multi-client eye tracking research is still impossible without inter-computer communication. We present a module that serves as a central hub and controller, which allows researchers to easily extend eye-tracking studies to many participants. We show how these two modules were used in an iterated public goods game. This application also shows the ease of integrating the modules with PsychoPy, a commonly used presentation library. We see important applications for this technology in behavioral economics, collective behavior, and other fields studying social interactions.

Corresponding author contact: dmedema@indiana.edu

11:05 Measuring multiscale temporal clustering in speech signals

Abney & Kello
University of California, Merced

Human speech signals exhibit nested clustering in acoustic onset events. This clustering may arise from nested levels of linguistic representation: Sentences are collections of syntactic phrases, phrases of words, words of syllables, syllables of phonemes, etc. We will describe and demonstrate the Allan Factor (AF) function to measure multiscale temporal clustering in speech signals. AF variance is a measure of temporal clustering in a point process at a given timescale, and the AF function computes AF variance across a range of timescales (e.g. milliseconds, seconds, minutes, etc.). A flat AF function means no clustering (i.e. a Poisson point process), and a scaling law in the AF function means multiscale clustering. In the presentation, we will proceed through the steps of the method by: (1) starting with an acoustic waveform of a speech signal, (2) converting the waveform into a time series of acoustic onset events akin to a spike train, and (3) computing the AF function. We will show aggregate AF results for a few different speech datasets, and we will demonstrate how the AF method can be extended to study human interaction and conversation by comparing AF functions between two interlocutors.

Corresponding author contact: dabney@ucmerced.edu

11:20 Voice and gesture in schizophrenia: A machine learning approach.

Simonsen & Fusaroli
Aarhus University

Contrary to other health impairments, psychiatric diagnosis and monitoring still widely relies on practitioners' qualitative evaluation (e.g. structured interviews). This process is time-consuming and expensive for patients and the medical system. In this study we recorded voice and gesture data during diagnostic interviewing of 37 participants with schizophrenia and 22 matched controls. We automatically extracted patterns of speech rate, pause distribution and wrist acceleration, calculating mean, standard deviation and recurrence quantification analysis over time. We used 5-fold cross-validated ElasticNet feature selection and logistic regression to calculate the diagnosis, accuracy balanced via Bayesian variational inference. Speech rate and pause indexes led to a 80.66% (CI: 72.04; 87.91) accurate reconstruction of the diagnosis. Particularly informative were variations in recurrence rate and mean of speech rate and number of pauses, all higher for the participants with schizophrenia. Gesture indexes led to a balanced accuracy of 83.58% (CI: 74.74; 90.96). Particularly indicative were: reduced and less varied movement in the participants with schizophrenia, accompanied by more varied and structured movement in the interviewer. Gestural indexes from the interviewer displayed the same amount of information as gestural indexes from the participants (accuracy ca. 80%). Employing both gesture and voice led to a balanced accuracy of 88.61% (CI: 80.89; 94.56). The findings suggest computer-based unobtrusive and inexpensive ways of supporting the diagnostic and monitoring processes in schizophrenia and other mental disorders.

Corresponding author contact: arndis.simonsen@clin.au.dk

11:35 User-input without user-input: Quantifying spontaneous cognitive processing in listening-type tasks.

Huette & Viaud
University of Memphis

Ecological validity in testing theories of cognition affects the ability to predict and generalize, as well as putting theory to practice with applied research. Many times an explicit task can interrupt processing in non-naturalistic ways. In previous research, Huette et al., (2014) demonstrated an effect of grammatical aspect on eyemovements to a blank screen without any overt responding or awareness that attention was required. This type of cross-modal influence of language on vision without the need for user engagement supports the idea that linguistically delivered information can mediate the way people look for information in their environment. In the current work, negation was used in conjunction with pictures of everyday visual scenes to demonstrate negation functioning as a contextual signal to search an environment for possible alternatives. For example, “John did not eat the eggs” is presented before a picture with a plate of eggs, potatoes, and toast. In all scenarios, participants were predicted to look around the screen more in the negated language condition, compared to an affirmative language condition. This general increase in visually searching the screen when negation is present is shown with a number of different overall metrics including difference from affirmative language, difference from what bottom-up visual saliency would predict, and measurements relating to the distribution of eye movements including kurtosis and area of the convex hull. These measurements have the ability to look at what happens to a participant “spontaneously” without interrupting cognitive or perceptual processing by asking for explicit, overt responses.

Corresponding author contact: shuette@memphis.edu

11:50 Monitoring behavior in a game-based tutoring system.

Russell, Snow, Allen & McNamara
Arizona State University

Students’ ability to self-regulate is a critical component of the learning process. One key element of self-regulated learning (SRL) is monitoring, or the cognitive process that assesses states of progress relative to goals and generates feedback to guide further action (Butler & Winne, 1995). However, students’ often struggle with this skill. According to one view, problems with monitoring arise not because the cognitive process of monitoring is failing, but simply because information was or was not monitored (Winne, 1989). It is our hypothesis that increasing the ways in which students can monitor their progress may combat poor monitoring behavior. The current research explores how the usage of monitoring features in an Intelligent Tutoring System (ITS) impacts user performance.

University students (n = 75) interacted with the ITS iSTART-2 over a 3-hour session. During the game-based practice phase of the experiment, participants were exposed to the game-based menu and were free to interact with the system interface in any order. Participants could also check the quality of the self-explanations they had generated, and view trophies they had earned from the games (monitoring behaviors). Correlation and regression analyses were calculated between measures of self-monitoring (frequency of checking self-explanation quality or trophies) and ITS performance (posttest self-explanation quality). Results indicated that monitoring behaviors

significantly predicted variance in performance, and that users who more frequently engaged in instances of self-monitoring generated better self-explanations.

Corresponding author contact: devin.russell@asu.edu

12:05 Moved to poster session

8:00 Does the Competition-based Conversational Agents Promote Reading?

Li, Cai & Graesser
University of Memphis

Previous research showed game-based elements in intelligent tutoring systems (ITS) motivated reading and yielded better reading performance. This study further investigated whether the competitive agents combined with the game-like elements such as visible feedback and competition-based scoreboard designs would motivate adult readers and enhance their reading performance. In this study, 50 college students participated in the conversational dialogues study—a teacher agent, a student agent, and a human student—in the adaptive system. They were asked to identify the appropriate meaning of the words with multiple meanings in the context. There were 30 items at three difficulty levels in total: 10 at easy level, 10 medium, and 10 difficult. Each student finished the total 9/10 items from different difficulty levels based on the amount of correct items. For each item, the sentence was displayed at the top of the screen, with the target word underlined, followed by two images related to two different meanings of the target words in the middle of the screen. The teacher agent asked the “two students” to choose the correct meaning and then provided the immediate verbal and visible feedback. Two conditions were designed after the teacher’s feedback: displaying the scores in the scoreboard (Scoreboard) in screen or not (Non-scoreboard). After doing the activity, the students were asked to take the survey about the motivation and engagement with 1-6 Likert scale. The mixed modeling results showed scoreboard condition yielded a slightly higher reading performance, and felt more motivated and engaged than nonscoreboard condition.

Corresponding author: graesser@memphis.edu

8:15 Developing reading comprehension lessons in AutoTutor for The center for the study of adult literacy

Baer, Hays, Cai, & Graesser
Institute for Intelligent Systems, University of Memphis

The Center for the Study of Adult Literacy (CSAL) is committed to understanding the reading-related characteristics that are critical to helping adult learners (readers at the 3rd-8th grade levels) reach their reading goals and to developing instructional approaches tailored to their needs and interests. Adults who struggle with reading are a heterogeneous group with an extremely varied set of skills and experiences. Difficult circumstances often dictate their ability to attend classes regularly. Adopting a web-based instructional tutor allows for individualization of instruction, and increases engagement in learning experiences inside and outside of the classroom. Our web-based series will involve animated conversational agents, a curriculum script, semantic evaluation of student contributions, and adaptive conversational dialogues. Each lesson uses the AutoTutor framework to deliver a review video of the didactic instruction, exercises based in practical topics to assess understanding, and independent reading using web-based texts from our repository (which includes over 6000 web-based texts leveled in Coh-

Metrix). One of the challenges in designing the lessons is determining the level of complexity in computer interaction that is suitable for the population. Due to the extreme variation in learner's prior knowledge, we want to provide an experience that introduces the learner to new digital skills while emphasizing the comprehension strategies that are the focus of our lessons. Some learners are apprehensive about the skills required to use the interface, but we have found that they are intrigued by the themes of the content and the support they receive from guided practice with the agents.

Corresponding author: wbaer@memphis.edu

8:30 AutoTutor Framework for Integrating Conversations into Intelligent Tutoring Systems

Cai, Graesser, Hu & Nye
University of Memphis

AutoTutor is an established framework that supports conversions for intelligent systems, with an emphasis on intelligent tutoring systems. The AutoTutor framework is based on extensive research on tutoring, natural language processing and discourse analysis. Recent advances enable AutoTutor to more widely integrate with other systems. The major components of AutoTutor framework include ACE (AutoTutor Conversation Engine), ASAT (AutoTutor Script Authoring Tool), and a set of AutoTutor delivery interfaces. ACE interprets AutoTutor scripts containing conversation transition rules and conversation content. ACE is responsible for assessing the quality of user responses, interpreting conversation transition rules, and sending the client program action sequences to execute. ACE also provides detailed interaction logs which can be used by the client to make immediate tutoring decisions or to record for researchers studying the conversation sessions later. Conversational scripts for ACE are designed using ASAT, which has several different forms. ASAT-D is a desktop version that is most powerful and flexible, supporting the design of challenging dialog rules. This version is primarily used by specially-trained authors. ASAT-W is a web version. While it provides limited conversational rule authoring, it facilitates online collaborative authoring of simpler conversation scripts. ASAT-V is a visual version of ASAT where conversational rules are designed using flow charts, which greatly reduces the difficulty of conversation rule design. AutoTutor client interfaces exist as a component in ASAT and also as standalone web applications. Through this set of tools, AutoTutor framework makes it possible to integrate intelligent conversations into a variety of systems.

Corresponding author contact: zcaai@memphis.edu

8:45 Gains in metacognitive awareness and writing proficiency in an intelligent tutoring system for writing

Allen, Roscoe, Snow, & McNamara
Learning Sciences Institute, Department of Psychology, Arizona State University

National assessments demonstrate that many students write below proficiency. To remediate this problem, researchers have developed automated writing evaluation systems that provide students with instruction, practice, and feedback on their writing. However, despite widespread implementation of such systems, it remains unclear how students respond to automated feedback and, ultimately, whether students benefit in terms of metacognitive awareness or writing

proficiency. The current study examined changes in both the accuracy of students' self-assessments and the quality of their writing across eight sessions with a writing strategy intelligent tutoring system (ITS). Additionally, we investigated whether specific individual differences were more or less predictive of observed performance gains (i.e., increases in self-assessment accuracy and essay scores). High school students ($n = 88$) completed eight training sessions in an ITS designed to promote writing proficiency. During each session, students wrote an essay and provided a self-assessment of essay quality. The system subsequently provided both summative and formative feedback on these essays. Results indicated that, on average, training with the ITS led to significant increases in writing performance and self-assessment accuracy. Additionally, gains were related to key individual differences including motivation and prior knowledge. Overall, these results suggest that automated writing evaluation systems can promote increases in students' ability to produce higher quality essays along with heightened awareness of their own performance. Importantly, these improvements may be stronger for students who are more motivated to learn and students who possess more prior knowledge.

Corresponding author contact: laurakallen@asu.edu

9:00 Does agency matter?: Path analysis within a game-based system

Snow, Allen, Jacovina, & McNamara
Arizona State University

When students exhibit control and employ a strategic plan of action over a situation they are said to be demonstrating *agency* (Bandura, 2001). The current work is comprised of two studies designed to investigate how agency manifests within students' choice patterns and ultimately influences self-explanation quality within the game-based system iSTART-2. In Study 1, 75 college students interacted freely within iSTART-2 for two hours. Using random walks and Entropy analyses, we visualized and quantified students' choice patterns as either controlled or random and examined how this related to self-explanation performance within iSTART-2. Findings revealed that students who demonstrated controlled choice patterns generated higher-quality self-explanations compared to students who exhibited random choice patterns. We hypothesized that the advantage for students who exhibited a controlled pattern was in part driven by their experience of agency. That is, engaging in controlled patterns should be advantageous only when doing so is a result of students' strategic planning. In Study 2, we tested this hypothesis by assigning 69 students a choice pattern (i.e., controlled or random) that had been yoked to students from Study 1, thus removing students' ability to exert agency over their learning. Findings from this study revealed no differences in self-explanation quality between the groups assigned to random and controlled choice patterns. Collectively, findings from these studies support the notion that success within game-based systems is related to students' ability to exert agency over their learning paths.

Corresponding author contact: Erica.L.Snow@asu.edu

Leveraging Online and Network Tools

9:30-10:45 am

Seaview Ballroom C

9:30 Cutting the Cord: Capitalizing on Google Glass for Naturalistic Interaction Research

Rodriguez, Paxton & Dale
University of California, Merced

Recent advances in wearable technology are extending opportunities to explore cognitive phenomena in naturalistic settings at a lower cost than ever before. Google Glass -- glasses-like wearable technology recently released by Google -- is perhaps the most publicized of these devices. Glass's wireless computing and sensing technologies allow researchers to track complex behavior while moving outside the laboratory setting. We here present our recent efforts to prepare Google Glass for interpersonal interaction research and provide guidance for others interested in utilizing Glass.

Corresponding author contact: krdriguez28@ucmerced.edu

9:45 A new HTML5 library for online data collection

Travers
Queen's University Belfast

PsychScript is a set of HTML5/JavaScript tools for conducting psychological experiments over the internet. It consists of a library of JavaScript functions, HTML and CSS templates, and a simple PHP/SQL routine for storing participants' data on a server. PsychScript provides a low-level API, which is easier to learn and use than pure JavaScript, but provides greater flexibility and control than more high-level tools such as Qualtrix, Inquisit, SurveyMonkey, and the jsPsych JavaScript library. Advanced stimuli presentation and user input, including the mouse-tracking paradigm, can be achieved using the HTML5 canvas element, in conjunction with PsychScript's built-in functions. Development of PsychScript is completely open source, and source code is available on GitHub. This presentation will demonstrate how popular paradigms can be implemented in PsychScript, and discuss the state of the art and technical issues faced in online research.

Corresponding author contact: etravers01@qub.ac.uk

10:00 NeuroVirtual 3D: A Virtual Reality platform for psychology

Cipresso*, Serino* & Riva*[^]
*Istituto Auxologico Italiano, Milano, Italy
[^] Università Cattolica, Milano, Italy

Virtual Reality has been extensively used in a wide range of psychological experiments. With this communication we aimed at presenting NeuroVirtual 3D, a platform expressly thought to be used by psychologists. It rely on the experience developed with NeuroVR software, but hugely extended for experiments. The software is made available for free to Researchers, which can also use a large number of virtual environments and objects already developed. The platform has

been developed to connect to virtually every device ever produced by the means of VRPN protocols, however a number of these have already been included and tested in the platform. Among the available device has already been configured the Microsoft Kinect low-cost sensor for navigation through the virtual environments and to trigger specific action (sounds, videos, images, ...) when a specific gesture is recognized, e.g. a step forward or an arm up. In the platform has already been implemented a task for Neglect and a task for spatial abilities assessment. More, NeuroVirtual 3D integrated a TCP-IP-based module (bridge) to collect data from virtually any existent biosensor (Thought-Technology, Zephyr and StarStim devices have already been included in the platform). It is possible to record any psychophysiological signal during any experiment, also using the computed indexes in realtime. More in general NeuroVirtual 3D is able to record external and internal (e.g., coordinates, keys-press, timestamp) data with a millisecond precision, representing de facto the most advanced technology for experimental psychology using virtual environments, at date available without the needs to program code.

Corresponding author contact: p.cipresso@auxologico.it

10:15 Effective communication of risks for Android apps: Influence of summary risk information and framing

Chen, Gates, Li, & Proctor
Purdue University

Due to its openness and popularity, the Android platform has become a target for malware. The current way of conveying potential risks for the apps is to provide users with a list of the permissions that an app requests. However, this method relies heavily on users to comprehend the permissions and to make appropriate decisions, which has been shown to be ineffective. We proposed a solution, which is to present a summary index of risk/safety information, and have obtained evidence that this index helps users to select apps with lower risk or higher safety, especially when the index is framed in terms of safety. We summarize that evidence in this talk. In one study, we conducted both online experiments through Amazon's Mechanical Turk and laboratory experiments, and a summary risk index, either in the format of texts or symbols, was added to the default interface in the Google Play store. Participants found this index to be useful and took it into consideration in their app selection decisions. In Study 2, another more ecologically valid online experiment further compared the presentation of this summary index in terms of risk and safety. Participants showed better comprehension of the index, and it led to selection of less risky apps when framed in terms of safety rather than risk. To conclude, our studies show that a summary risk/safety index is effective in promoting less risky or safer decision making in Android app selection and that this index is more effective when framed as safety.

Corresponding author contact: chen548@purdue.edu

Poster Session**1:30-2:30 pm****room TBA**

Poster Presenters: Please follow the number designation for your poster setup. Poster must be set up between 11:30-12:30 pm and must be removed by 2:00 pm. *Presenters of odd-numbered posters are to be at their posters for the first 30 minutes, and presenters of even-numbered posters at theirs for the second 30 minutes of the poster session.*

Educational Systems

1. Li, Graesser, & Shaffer - What Is Associated with Question Asking in Epistemic Games
2. Cai, Li, Graesser & Hu - ASAT-V: A Visualized Authoring Tool for AutoTutor Conversations
3. Koch & Shinseki - A Student-Based Evaluation of PSPP
4. Gomez & Allbritton - A Concept Inventory for Elementary Statistics
5. Fang, Hu & Nye - Detecting Motivation in Online Mathematics Learning

Cognitive Processes

6. Redick, McAbee, Oswald & Hambrick - Validation of a Short Domain-general Measure of Working Memory Capacity
7. Shono, Koshino & Stacy - Novel Lexical Decision Task Using Phrasal Primes
8. Guo, Trueblood & Diederich - An Investigation of Time Pressure on Framing Effects in Risky Choice
9. Liao, Wei & Lin - Effects of stimulus codes and response modalities on spatial Stroop interference
10. Salamanca, Schultz & Washburn - The Attrition Problem with Transcranial Doppler Sonography

Modeling and Semantic Analysis

11. Cunningham, Burgess & Estep - A Self-contained Multi-dimensional Scaling Program Designed for the Presentation of High-dimensional Memory Modeling Results
12. Lippert - The Psychology of the Supreme Court: Modeling Judicial Semantics from Written Opinions
13. Samei & Luno - What is it (not) about?: Adjusting Frequency Measures for Genre Classification
14. Bermudez, Carroll, Burgess & Estop - Using Semantic Distance Congruency to Predict Human Ratings of Gender- Occupation Bias

15. Cao & Burgess - The Connotation of the N-word in Twain's Novels and a Comparison of Models Trained on Usenet Text (1994 and 2005)
16. Huang, Shubeck, Xie, Cai, Nye, & Hu - Identifying Heuristics for the Effect of Parameter Changes on Semantic Space Performance
17. Xie, Huang, Wang, Cai, Nye & Hu - Evaluating domain glossaries for domain-specific topic evolution in Sina Weibo.

Tools and Issues

18. Markant, Chan, Coenen, Martic, McDonnell, Rich & Gureckis - PsiTurk: A framework for running online behavioral experiments
19. Stacy, Shono & Ames - Utility of Computerized Assessment in Field Settings and At-Risk Group
20. Zuccolotto, Babjack, Cernicky, Sobotka, Basler & Struthers - Methods for assessing and standardizing audio stimulus presentation latencies across heterogeneous hardware and operating system platforms
21. Bain, Weishaar, Kilic & Gomez - A novel method for displaying behavioral data: Plotting accuracy against vincentiles of reaction time.
22. Lindemann & Krause - Developing Cognitive and Neuroscientific Experiments for Tablet Computers and Mobile Phones under Python: The Expyriment Android Runtime
23. Roux, Baene, & Carreiras – A framework for the automated analysis of speech data
24. Bannert & Reips - LogR: An R package to analyze log files
25. Brown & Hays - Effect of a post-paid incentive on response rates to a web-based survey

Poster Abstracts

Educational Systems

Does the Competition-based Conversational Agents Promote Reading?

Li* Graesser* Shaffer**

*University of Memphis

** University of Wisconsin-Madison

Question asking has been investigated in epistemic games to study the problem solving skills of the students. For example, in epistemic games, the frequency of questions increased in task difficulty and task unfamiliarity, and students tended to ask more questions than the mentor per hour. However, little research was conducted on whether students and program formats are associated with the question asking. This study investigated whether the students' grade level, and the type of school and the format of the game they attended were associated with the questions they asked in an epistemic game, Land Science. Land Science is a multi-party game to simulate urban-planning internship. A mentor assigned the tasks and guided the students to finish their tasks. All the students could communicate with others through text chats for inquiries. Participants were categorized into four groups based on their grade levels: 10th graders, 8-12th graders, 9-12th graders and AP students. Meanwhile, they could be grouped into 2 types, regular school students and home school students. Finally, based on the formats they attended the program, they could be classified into vacation program and in-school program. The frequency of questions per hour per student was analyzed with the mixed modeling. Results showed that frequency of question that students asked was different in different graders, but not in school types or in program format. Further study was conducted to examine the question distribution of each student in the game, which may help interpret why particular groups generated more questions than others.

Corresponding author contact: hli5@memphis.edu

ASAT-V: A Visualized Authoring Tool for AutoTutor Conversations

Cai, Li, Graesser & Hu

University of Memphis

AutoTutor conversations have been used in many intelligent tutoring, mentoring, and assessment systems. The form-based rule design tool turned out to be powerful, flexible but hard to use. ASAT-V, a flowchart-based authoring tool, visualizes the complex AutoTutor conversation rules. While the flowchart can be generated by any drawing tool with necessary functionalities, the current version of ASAT-V only accepts flowchart generated by Visio 2013. In a Visio stencil, we defined all AutoTutor shapes, including *Start*, *End*, *Agent*, *Speech*, *Question*, *Answer*, *Event*, *Action*, *Transition* and *Connector*. Each AutoTutor has a specific set of properties. For example, the properties for *Speech* shape include "Name," "Agent," "Text," and "Speech." One conversation script can be put on multiple Visio pages, connected by "Transition" shape. Authors use Visio 2013 to create AutoTutor scripts by dragging and dropping the shapes, specifying the properties and making connections. After a flowchart is composed, authors use ASAT-V to check the validity of the flowcharts and further define agents and common speech acts. There is a test panel in ASAT-V for authors to simulate the conversations and test whether the conversations work as the authors expected. Authors can easily switch between ASAT-V and Visio 2013. Once finished, the flowchart file can be uploaded to AutoTutor servers, where ACE (AutoTutor

Conversation Engine) provides an interpreting service which can be consumed by any user interface program through JavaScript function calls. ASAT-V makes it possible for domain experts to efficiently create high quality AutoTutor conversation scripts.

Corresponding author contact: zcai@memphis.edu

A Student-Based Evaluation of PSPP

Koch & Shinseki
George Fox University

GNU PSPP version 0.8.3 is a statistical software package developed through the Free Software Foundation. It is designed to resemble, function like, and replace SPSS. The syntax and data files are compatible with SPSS and PSPP opens SPSS files. Given the non-propriety availability of PSPP, it may be a useful tool for teaching students how to conduct statistical analyses. A class of 34 graduate students in a professional program in the Pacific Northwest used PSPP and Excel during the course of a semester and then evaluated both programs after completing a professional practicum. Twenty-one students completed the evaluation. Specific to PSPP, students evaluated the software regarding the ease of downloading and installing, documentation, and aspects of usability. Ratings were generally noncommittal, however, students typically believed that the documentation was lacking. Students were asked parallel questions about using PSPP and Excel. Specifically, they were asked to indicate which was more difficult: knowing what analysis to do, knowing how to do the analysis with PSPP/Excel, or if they were both equally difficult. Students were more likely to indicate that “knowing how to use PSPP” was the most difficult ($X^2(2) = 7.6, p < .05$) but were equally likely to select “knowing how to do the analysis” and “both were equally difficult” for Excel. Statistical self-efficacy was positively correlated with PSPP documentation and icon ratings along with applying concepts from the course in the practicum experience. Overall, PSPP may be useful for teaching statistics but its effectiveness may be moderated by other factors. [250 words]

Corresponding author contact: ckoch@georgefox.edu

A Concept Inventory for Elementary Statistics

Gomez & Allbritton
University of DePaul

A concept inventory, the DePaul Statistics Inventory (DPSI), is presented for assessing conceptual understanding of elementary statistics. A concept inventory is a criterion-referenced test designed to evaluate working knowledge of a specific set of concepts. Perhaps the best known example is the Force Concept Inventory (FCI) (Halloun & Hestenes, 1985) used to identify misconceptions about mechanics. A surprising finding from assessments using the FCI was that the gains after taking a course in Physics were rather modest and only marginally related to the students' grades. We wished to know whether students taking statistics courses in psychology would demonstrate an increase in conceptual understanding of probability and statistics, as measured by the DPSI.

The DPSI is a multiple choice test of basic statistics concepts that does not require the use of formulas, equations, or statistical software. In contrast to previous statistics inventories (e.g., Allen, 2006, PhD dissertation), the DPSI avoids the use of technical terms and does not depend on students' ability to remember definitions. Thus we were able to test students' intuitions before taking any courses in

statistics as well as after one or more courses. We present the results of an assessment using the DPSI, along with an R based explanation of all answers.

Corresponding author contact: pgomez1@depaul.edu

Detecting Motivation in Online Mathematics Learning

Fang, Hu & Nye
University of Memphis

In an online learning environment, unlike face-to-face teaching contexts where instructors are able to observe student motivation, it is difficult to detect student motivation. However, detecting student motivation is important, because less-motivated students run the risk of being “out-of-sight, out-of-mind” in an online environment. Previous studies found out that internal motivation was positively related to persistence, academic achievement, in-depth learning, and preferences for optimal levels of challenge. In the current state-of-the-art, intelligent tutoring systems supported with sensors (e.g., facial expression classifiers) and specialized human-computer interaction (HCI) interfaces have been used to detect and react to varying motivational states. Unfortunately, this type of dedicated instrumentation is often expensive and not easily integrated into authentic learning contexts. This work attempts to detect student motivation by analyzing moment-to-moment learning behavior that is shown in the time sequence data provided by the system. It builds on our preliminary online mathematics study in which the teacher who monitored the classes evaluated student motivational level (including whether students were motivated to learn and whether they volunteered to math class and more responsibilities) and recorded these estimates. It was found that student motivation, combined with prior knowledge, could predict students’ academic achievements. In this study, we are trying to study the learning patterns of students with different motivation levels. We will measure two constructs of motivation—perceived competence and tension, with a questionnaire. After the measurement, we will divide the students into several groups based on their motivation scores. By analyzing the time sequence data of different groups of students, we will search for distinct learning patterns between these different groups. If distinctive patterns are shown, we will identify features that can be used as the indicators of motivation.

Corresponding author contact: xiangenu@gmail.com

Cognitive Processes

Validation of a Short Domain-general Measure of Working Memory Capacity

Redick*, McAbee**, Oswald** & Hambrick****

*Purdue University

** Rice University

*** Michigan State University

Working memory capacity is among the most frequently assessed constructs within psychology. Complex span tasks, which involve a processing component that serves as a distractor from the ongoing serial recall memory component, are measures of individual differences of working memory capacity that have demonstrated desirable psychometric properties. Recent advances to fully automate the complex span tasks and provide them for free download have been beneficial for researchers. However, one practical issue limiting the widespread use of complex span tasks in applied and lab research is that administering multiple tests to measure verbal and spatial working memory capacity is time-consuming,

which creates a conflict between managing limited testing time and measuring numerous constructs reliably. Previously, using a series of confirmatory factor analysis models on a large archival data set of three complex span tasks, we developed a short, domain-general working memory capacity measure following a principled psychometric process. Using the measure properties specified in this previous work, we collected data from N = 185 subjects on the shortened measures, and also administered fluid intelligence tests to examine the shortened measures' criterion-related validity. The shortened measures demonstrated acceptable levels of reliability, and a confirmatory factor analysis showed a strong relationship to a fluid intelligence latent variable, consistent with previous work with the full versions of the complex span tasks. The shortened measures also reduced the duration of the experimental trials by over 50%. Therefore, the shortened complex span tasks measure domain-general working memory capacity quickly without drastically altering validity or reliability.

Corresponding author contact: tredick@purdue.edu

Novel Lexical Decision Task Using Phrasal Primes

Shono*, Stacy* & Koshino**

* Claremont Graduate University

** California State University, San Bernardino

The present study conducted an associative priming experiment with a novel lexical decision task (LDT) that used two-word phrases as primes. In a typical priming experiment that uses a visual word LDT, primes often consist of either a single word (e.g., “blue”) or multiple single words (e.g., “blue”, “cloud”). Consistent results have shown that processing of a target (e.g., “SKY”) is facilitated when it was preceded by an associated prime (e.g., “blue”), relative to when preceded by an unassociated prime (e.g., “desk”). However, there is evidence that short phrases, rather than words, could also serve as useful elements of associative memory as revealed by free association tasks (e.g., Stacy, 1997; Stacy, Leigh, & Weingardt, 1994). The current experiment examined if the novel LDT with phrasal primes would produce comparable associative priming effects often obtained in typical LDTs. Based on associative prime-target pairs (e.g., “blue – SKY”) used in past priming experiments, we created two-word phrasal primes while keeping their original associative relations between the primes and targets (e.g., “clear blue – SKY”). In the LDT, two-word phrasal primes were presented one at a time (133 ms each), followed by a blank (33 ms) and a target. Thus, SOA was kept at 300 ms to ensure that automatic associative processes were measured. Results showed that associated phrasal primes produced a significant associative priming effect as compared to unassociated phrasal primes (mean priming effect = 29.00 ms). Implications of the findings are discussed from an associative memory network viewpoint.

Corresponding author contact: yusuke.shono@cgu.edu

An Investigation of Time Pressure on Framing Effects in Risky Choice

Guo*, Trueblood* & Diederich **

* University of California, Irvine

** Jacobs University

This work investigates the intuitive and deliberative thinking modes in risky decision-making by manipulating time pressure. A recent fMRI study by De Martino et al. (2006) found greater activation of the amygdala when exhibiting framing effects (risk-averse behavior in gains and risk-seeking behavior in losses), suggesting that they may be driven by an intuitive / emotional system. Based on this

hypothesis, we developed a new experiment to examine the effect of time pressure on framing effects. If framing effects are the result of a fast, emotional system, then we expect more pronounced framing effects under time pressure. In our experiments, participants made choices between a gamble and a sure option, which were either framed as gains or losses. We also included a speed manipulation in which time pressure is applied to participants to make decisions quickly and an accuracy manipulation in which participants are encouraged to take as long as necessary to maximize the amount of money earned. Choice probabilities and response times are modeled by a diffusion model in which the quick, intuitive system activates first, after which a switch occurs and the thinking mode is shifted to the deliberative system which has a different drift rate. We explore models in which the drift rates and boundaries vary in accordance with the thinking modes, frames, and time pressure.

Corresponding author contact: a.diederich@jacobs-university.de

Effects of stimulus codes and response modalities on spatial Stroop interference

Liao, Wei & Lin
National Chung Cheng University

The present study investigated the spatial Stroop effect under different pairings of stimulus codes and response modalities. Stimulus code (word vs. arrow), response modality (keypress vs. mouse movement vs. touchscreen tapping vs. touchscreen dragging), directional information (left vs. right) and physical location of the stimulus code (center vs. left vs. right) were manipulated. Participants' task was to follow the directional information of the stimulus to press left or right keypad, tap the left or right box, or move a cursor to the left or right box on the touchscreen by dragging or mouse movement. Spatial Stroop interference effect occurred when the directional information of the stimulus was incongruent with its physical location. Results showed greater reaction time for the word than for the arrow for all response modalities. The spatial Stroop interference was observed in participants' greater reaction time for the keypress, touchscreen tapping and dragging responses. For the mouse response, greater spatial Stroop interference was observed in movement time, path length and direction changes during the movement for the word than for the arrow. Comparing to touchscreen and keypress responses, mouse movements demanded more attentional resources in monitoring the cursor movement and thus was more easily affected by the less compatible word stimulus. Although dragging response also involved cursor movements, less resource was demanded from direct manipulation and so no effect was found for the word stimulus. The present results may provide suggestions for the selection of input device for an interface display when spatial interference is of concerned.

Corresponding author contact: mjliao@ccu.edu.tw

The Attrition Problem with Transcranial Doppler Sonography

Salamanca, Schultz & Washburn
Georgia State University

Transcranial Doppler (TCD) sonography is a computer-based neuroimaging technique in which changes in cerebral bloodflow velocity related to neuronal activity are measured noninvasively via ultrasound. Transducers record bilateral changes in the speed of blood perfusion that result from alterations in mental workload. Although TCD has limited spatial resolution, it affords numerous benefits relative to other imaging paradigms, including high temporal resolution, relatively low cost, rapid response rate, and relative comfort and ease of use. However, TCD does have an attrition problem : for about half of

the participants one tests, the cerebral arteries in both hemispheres can be reliably sonated; but for the remaining participants, one cannot obtain a valid signal for the left hemisphere or the right hemisphere or both. The question for the present analysis is whether any interpretation can be made of this pattern of effects. That is, are participants for whom bilateral signals can be identified different with respect to cognitive performance, cerebral bloodflow, or patterns of changes in either, compared to participants from whom only a unilateral signal can be obtained? We compared these three groups (Both, Left only, Right only) across measures. Although no differences between the groups were found in baseline bloodflow velocity, there were Task X Hemisphere differences, reflecting functional cerebral asymmetries that could only be captured with bilateral recording. We attempted, with limited success, to develop an algorithm for estimating bloodflow velocity in the unmeasured hemisphere, given available data from the recorded cerebral artery.

Corresponding author contact: dwashburn@gsu.edu

Modeling and Semantic Analysis

A Self-contained Multi-dimensional Scaling Program Designed for the Presentation of High-dimensional Memory Modeling Results

Cunningham*, Burgess* & Estep**

* University of California, Riverside

** San Diego Mesa College

Data visualization is crucial in presenting the results of high-dimensional memory and language modeling. Although there is no shortage of software that can be used for multi-dimensional scaling, what we present is a program that can accept co-occurrence vectors in various formats to accommodate a range of existing modeling software. The program runs as an application from the Mac App Store on OSX. Since MDS does not require linear relationships it is ideal for the visual representation of the dimensionality reduction that is necessary to produce a figure suitable for presentation or publication. The program can maintain the vector labels that can be important in conveying semantic or grammatical relationships. The program includes a mechanism to check the reliability of the MDS solution to deal with the problem that MDS ordinations are not unique solutions and local minima needs to be avoided. The program offers a solution to potential users of high-dimensional memory models who may not be well versed in the more sophisticated software products such as Matlab or R.

Corresponding author contact: acunn001@ucr.edu

The Psychology of the Supreme Court: Modeling Judicial Semantics from Written Opinions

Lippert

University of Kentucky

The construction of knowledge networks from text is a novel way to study the cognitive organization of domain specific content. This work evaluates the application of knowledge networks to legal text. Text analysis methods were used to transform text from 8,014 Supreme Court opinions into matrix data suitable for the construction of knowledge networks known as SCOD networks (Supreme Court Opinion Derived networks). Four specific hypotheses were then tested to better understand the meaningfulness and validity of SCOD networks. The first hypothesis considered differences between SCOD networks and random networks. The remaining hypotheses considered the ability of SCOD networks to reflect

known issues of the Court. Monte Carlo simulations, various graph theoretic measures and measures of graph similarity were used to test these hypotheses. Results showed significant structural differences between SCOD networks and random networks. SCOD networks were also shown to have good face validity in representing scholarly characterizations of the Supreme Court, and in particular reflected known issues concerning the influence of ideology on Supreme Court decision making. In general, this work demonstrates the potential in using knowledge networks derived from text to help answer a wide variety of questions concerning Supreme Court decision making and the nature of judicial expertise.

Corresponding author contact: annelipp@gmail.com

What is it (not) about?: Adjusting Frequency Measures for Genre Classification

Samei & Luno
University of Memphis

Computational linguistic measures used for text analysis (e.g. genre classification) are often based on lists of words predefined and validated by linguistic experts to represent a certain “kind of” language. These models are often called “bag of words” models. One such measuring system is LIWC (Pennebaker et al., 2007) in which the normalized frequencies of the word set (i.e. categories) is captured as a measure representing to what extent a document is linked to a category. One issue with such measures is that their frequency output is defined by presence of the words as a whole, not the coverage of a category, as the absence of certain words from the category is unconsidered. In this work we examine the potential issues with the traditional frequency measures and word lists, by defining mathematical formulation to take into account the coverage a given category demonstrates, in order to better assess the similarities between documents these frequencies intend to highlight, as well as determine the relative strength of categories as represented across text modes. The Lancaster-Oslo-Bergen, and the London-Lund Corpora used in Biber’s (1988) landmark genre classification study were utilized to investigate the performance of a genre classification model based on the proposed measures.

Corresponding author contact: bsamei@memphis.edu

Using Semantic Distance Congruency to Predict Human Ratings of Gender-Occupation Bias

Bermudez*, Carroll*, Burgess* & Estep**
*University of California, Riverside
** San Diego Mesa College

There has been considerable research on the social psychology of stereotypes and bias. The experimental methodologies employed in this research encompass traditional paper and pencil measures, explicit and implicit priming, and rating studies. In earlier research, we found that semantic representations from the HAL model were sensitive to racial bias in the media (Hatzopoulos, Crew, & Burgess, 2004) and ideological bias in political commentary (Maier, Burgess, Johnson, & Willits, 2005). Last year, we presented a study using HAL that investigated the racial connotation in the novel, *The Adventures of Tom Sawyer*. All these analyses had in common the comparison of word pairs that were either congruent or incongruent with the social bias (e.g., congruent examples: liberal-hope, conservative-business). In the present study we begin with the same process with gender bias and occupations. A set of female and male given names and a set of occupation titles were constructed and the distance was computed between all possible pairings (7,104) of names and occupations (e.g., Linda – secretary, John – contractor). We use the distances provided from HAL to predict human stereotype

judgments. Rather than determining if name-occupation pairings were congruent or not, subjects rated a set of occupations using their judgments of how masculine the occupation was on a 1 – 9 Likert scale. We evaluated both the city-block and the Euclidean distance metrics and whether or not standardizing the distances was important in the analyses. The key result to report is that the semantic distances reliably predicted the human gender stereotype ratings.

Corresponding author contact: lberm002@ucr.edu

The Connotation of the N-word in Twain’s Novels and a Comparison of Models Trained on Usenet Text (1994 and 2005).

Cao & Burgess
University of California, Riverside

At last year’s meeting, we (Burgess & Cao, 2013) presented the results of a semantic analysis of the n-word in Mark Twain’s novel *The Adventures of Tom Sawyer*. These novels, along with Twain’s *The Adventures of Huckleberry Finn*, have been constantly targeted for means of censorship from school libraries. There have been demands by school districts to remove the texts or to replace the novels with the n-word replaced with the word *slave*. Etymological research shows that during Twain’s era the N-word was relatively neutral in meaning and it was not until the late 1800’s and early 1900’s that the word acquired its pejorative and racist connotation. Our research with the HAL model confirmed that hypothesis. The model revealed that the distances for a set of adjectives with positive connotation were closer to the n-word than adjectives with a negative connotation. In the current study, we extended our analysis to the novel *The Adventures of Huckleberry Finn* with a similar analysis. We also conducted an analysis of the sentence contexts, as opposed to just the semantic word meaning. Last year we also showed how the semantic structure of the n-word was pervasively pejorative when the model was trained on Usenet text from 1994. In our second analysis, we compare the result from the model built from the 1994 text to a model built from Usenet text 10 years later.

Corresponding author contact: pcao002@ucr.edu

Evaluating domain glossaries for domain-specific topic evolution in Sina Weibo.

Xie*, Huang*, Wang***, Cai*, Nye* & Hu* **

* University of Memphis

** FNC Inc.

*** China Central Normal University

Tracking topics’ evolution within a micro-blog yields important insights into public opinion, which often intersects several domains (e.g. presidential elections have political, economic and military facets). One way to track a topic’s changes in related domains is to calculate the overlap between words in the topic’s discussion against related domain glossaries (i.e. words associated to a particular domain of knowledge). Existing glossaries using mainly formal and academic terms may not be suitable for informal posts that use a changing lexicon. Thus, new glossaries representing particular topics need to be created. However, we need to evaluate how much they accurately represent the domain of interest. This paper evaluates domain glossaries for a particular topic’s evolution in Sina Weibo, which is similar to Twitter and very popular in China. First, human raters classified topic-related posts into three domains, including social events ($Kappa=.949$), politics ($Kappa=.943$), and entertainment ($Kappa=.937$). Second, online news were collected to generate glossaries by assigning a word to a

domain if it had high entropy for only that domain. LSA vectors were constructed for each glossary and for sets of posts assigned to a single domain. These posts are clustered into three groups by K-means based on their semantic cosine dissimilarity. Next, each cluster is represented by an LSA vector and labeled with the domain having highest similarity with the cluster's vector. Assuming the glossaries reflect real domain differences, most of posts classified by computer should have the same domain as the human ratings.

Corresponding author contact: jxie2@memphis.edu

Tools and Issues

PsiTurk: A framework for running online behavioral experiments

Markant, Chan, Coenen, Martin, McDonnell, Rich & Gureckis
Max Planck Institute for Human Development

Abstract: Online experiments are growing in popularity in the behavioral sciences, offering a number of advantages including faster data collection and access to a more diverse pool of potential participants relative to lab-based research. However, a new set of technical and experimental challenges must be met before a researcher can adopt a web-based approach. psiTurk is a framework of software and web services that aims to address many of these challenges. It facilitates the creation, deployment, and sharing of web-based experiments that rely on Amazon Mechanical Turk (AMT), currently one of the most popular platforms for connecting participants to online experiments. The principal goal of psiTurk is to handle common technical challenges (including interfacing with AMT and managing participant pools) so that researchers can focus on the development and deployment of their experiments. Additionally, it offers a new online platform for scientists to share experimental code so that studies can be easily replicated and extended within the psiTurk framework.

Utility of Computerized Assessment in Field Settings and At-Risk Groups

Stacy, Shono & Ames
Claremont Graduate University

Computerized assessments of complex cognitive constructs are rarely administered in group settings in the field, ostensibly because of the realities and complexities in naturalistic or group situations where many potential participants can be found. Instead, assessment in group or naturalistic settings such as high-schools, community programs or workplace locations are usually restricted to paper and pencil surveys or to mere recruitment, for later individualized assessment either over computer, interview, or survey. The present research investigates the effectiveness of computerized assessment in settings that might be considered at best complex and at worst chaotic, such as in existing high-school classes and court-sanctioned drug education program classrooms. The study also examines the assessments in populations with risk profiles that might be assumed to preclude complex computerized assessment. Utility is examined in terms of psychometric properties, predictive utility or convergent validity, or patterns of missing data. Assessments include working memory, verb generation, other cognitive assessments, and survey measures. Study samples include students enrolled in alternative high schools because of academic, credit, behavior, or drug problems (N = 781) and adults enrolled in court-sanctioned drug education classes (N = 674). Participants were recruited and assessed in their existing, group settings, *in situ*. Results across samples demonstrated that most cognitive and survey measures showed adequate psychometrics and feasibility in the setting. Participants were amenable to the

assessments and completed comprehensive batteries within restricted time limits in complex, group settings in the field.

Corresponding author contact: alan.stacy@cgu.edu

Methods for assessing and standardizing audio stimulus presentation latencies across heterogeneous hardware and operating system platforms

Zuccolotto* Babjack*^ Cernicky* Sobotka* Basler* Struthers*

* Psychology Software Tools, Inc.

^ University of Pittsburgh

Audio stimulus presentation using heterogeneous computer platforms and audio output devices often introduces, 1) substantial variability across labs, and 2) variable time between intended sound delivery and actual sound delivery (sound onset latency). Fast and consistent audio onset latency is particularly important when audio stimuli need to be delivered precisely as part of EEG/ERP/psychophysical studies or in multi-site studies where standardization and strict control over the computer platforms used is not feasible. This paper introduces a novel approach to minimizing and assessing the latency and variability of audio presentation across different workstation configurations and processing applications. A stimulus presentation and latency assessment approach applicable to desktops/laptops/tablet computers is presented which uses E-Prime® and Chronos™ (a new multi-function, USB-based, data presentation and collection device). The present approach reliably delivers and/or timestamps audio stimuli with low latencies that vary by ≤ 1 msec independent of hardware and Windows OS/driver combinations. The Chronos™ audio subsystem adopts a buffering, aborting, querying, and re-mixing approach to the delivery of audio using E-Prime as the presentation system and standard E-Objects (e.g., Slide) to achieve consistent 1 msec audio onset latency without the use of advanced scripting. The resulting audio onset latencies are small, reliable, and consistent across systems. Additional features of the Chronos™ multi-function device are summarized which are useful in synchronizing and timestamping a wide range of stimulus presentation and response collection events.

Corresponding author contact: destiny.babjack@pstnet.com

A novel method for displaying behavioral data: Plotting accuracy against vincentiles of reaction time.

Bain, Weishaar, Kilic & Gomez

University of DePaul

When examining behavioral data, vincentile plots of reaction time (RT) yield important information about differences between conditions. However, vincentiles cannot be created from dichotomous data such as accuracy. A new graphing method for behavioral data was designed to accommodate accuracy. First, the data are separated by condition. Second, the midpoints between the vincentiles of interest are calculated for RT. For example, if the vincentiles of interest are the .1, .3, .5, .7, and .9 vincentiles, the .2, .4, .6, and .8 vincentiles need to be calculated for each condition. The data for each condition are then separated into groups for each vincentile of RT. The .1 vincentile group contains all data points below the .2 vincentile of RT, the .3 vincentile group contains all data points between the .2 and .4 vincentiles of RT, and so on. This process yields groups of data in which RTs cluster around each vincentile for each condition. Finally, mean accuracy is calculated for each of these groups. Points for each vincentile in each condition are graphed such that the horizontal position of each point is the RT for

that vincentile, and the vertical position is the accuracy for that vincentile group. Points are connected by lines that delineate the different conditions. This method displays differences in RT distributions between conditions, as well as the mean accuracy for the vincentile in each condition. The shapes of these lines have implications for modeling decision-making processes and differentiate between fast and slow errors.

Corresponding author contact: bbain3@depaul.edu

Developing Cognitive and Neuroscientific Experiments for Tablet Computers and Mobile Phones under Python: The Expyriment Android Runtime

Lindemann* & Krause**

* University of Potsdam, Germany

** Radboud University Nijmegen

The usage of tablet computers and mobile phones is becoming increasingly popular in psychology and neuroscience, as it offers the possibility to easily conduct empirical experiments outside the laboratory, assessing human cognition in everyday life situations. In contrast to classical experimental setups with desktop computers, very few tools exist to aid the development of experiments on touchscreen-based operating systems and mobile devices.

Here, we present a novel open-source runtime environment that allows researchers to conduct their experiments on mobile devices with the Android operation system. The experiment development is based on Expyriment, a free platform-independent lightweight Python library for timing-critical presentations and response assessment (Krause & Lindemann, 2014). The advantage of a Python-based development with Expyriment in combination with the Expyriment Android Runtime is that empirical studies originally conducted under highly controlled conditions inside the laboratory can be seamlessly transferred to a tablet computer or mobile phone with either no or only minimal changes in the source code.

Corresponding author contact: oliver.lindemann@uni-potsdam.de

A framework for the automated analysis of speech production data

Roux, Baene & Carreiras

Basque Center on Cognition

The analysis of speech production data requires a reliable way of measuring speech related features such as speech on- and off-set times and response accuracy. The assessment of these features, however, is often done manually through visual inspection of the data and as a result can be time consuming and influenced by subjective bias. Here, we present a framework for the automatic analysis of speech latencies and accuracy in speech production data. Our approach employs multitaper spectral analysis to extract a set of acoustic features in human speech from which speech latencies and accuracy are estimated. The performance of our approach was tested on $n = 10$ participants by quantifying the correlation between automatic scores and manually assessed response accuracies and response latencies. To demonstrate the suitability of our approach for the analysis of speech production data collected in the context of psycho-linguistic experiments, we will present data from a bilingual naming task where the present framework was used to automatically assess switch cost effects and to quantify the intra-individual stability of bilingual switch costs across different measurements

Corresponding author contact: f.roux@bcbl.eu

LogR: An R package to analyze log files

Bannert*,** & Reips*

*University of Konstanz, Germany

** ETH Zurich, Switzerland

Many scholars have recently moved to perform statistical computing and related tasks with R. R is a free software environment for statistical computing and graphics. It runs on a wide variety of UNIX platforms, Windows and MacOS. In the present paper we present LogR, a package written in R suited to analyze server log files. LogR offers seamless integration with other R packages and generally follows organizational criteria and computational algorithms suited to aid behavioral and social scientists as they were first developed in Scientific LogAnalyzer (<http://sclog.eu> , Reips & Stieger, 2004). LogR flexibly accepts a wide range of log file formats, various search options, computation of response times, detection of critical cases (e.g. possible multiple responses), speedy analysis, powerful visualization via a *shiny* web application server, and several common formats for export of data and reports. Several methodological features specifically needed in the analysis of data collected in Internet-based studies have been implemented in the R package and are described in this paper.

Effect of a post-paid incentive on response rates to a web-based survey

Brown* & Hays**

* RAND

** UCLA

We randomized half of the sample from two southern California medical centers to a post-paid incentive (n = 1,800) and half to no incentive (n = 1,800) for completing an internet survey about their experiences with health care. Respondents in the incentive group were given the choice between a \$5 cash or Target® e-certificate. The characteristics of respondents in the incentive and control groups was similar on age, education, length of membership in the plan, number of emails sent and visits to the primary care doctor in the 12 months prior to sampling, and their global rating of the doctor; the incentive group had more Asians (8% vs. 5%) and fewer Blacks/African Americans (2% vs. 5%) than the no incentive group.

Those randomized to the incentive were significantly more likely to respond to the survey than those in the control group (57% vs. 49%). Item non response rates did not differ between those in the incentive and the control groups. Those randomized to the incentive condition who completed the survey were significantly more likely to prefer a cash incentive over the e-certificate (70% of the incentives delivered to Internet respondents were in the form of cash). The unit cost per incentive was \$8.32 for cash and \$7.49 for the e-certificate. The results of this experiment indicate that a post-paid incentive can significantly increase the response rate to an internet survey.

NOTES:



Solutions for Research, Assessment, and Education

Providing resources to over 3,000 research institutions and laboratories throughout the world for 25 years.



Ask us about E-Prime training



Celeritas Fiber Optic Response System



E-Prime Extensions for fMRI

