



SCiP 2013

Society for Computers in Psychology

Toronto, November 14, 2013

Welcome Message from the President

Welcome to the 43rd annual meeting of the Society for Computers in Psychology (SCiP)! SCiP has a long tradition of advancing and disseminating leading work in core areas of psychological research pertaining to computers and computational techniques. This work includes methodology, instrumentation, statistical techniques and software toolboxes, in addition to more recent movements with neuroimaging techniques, cognitive modeling, educational applications, and web-based research.



It was exactly a decade ago that I first joined SCiP as a graduate student. I was instantly hooked on the high quality of fascinating research I learned about at the conference, and have faithfully attended every year since then. In the past decade I have also seen remarkable change in *how* psychological research is conducted, and SCiP is consistently leading the way with emerging techniques and discoveries. While the traditional strength of SCiP in the 70s and 80s was designing and using computational hardware for psychological experiments, SCiP in the 21st century is now synonymous with computational techniques for model testing, data mining, crowdsourcing, and clinical/educational applications.

One of the movements in science in general we are all becoming familiar with is the so-called “Big Data” trend. I feel that SCiP really has a leadership role in developing and guiding techniques, hardware, and models to advance theoretical understanding from large-scale data in addition to traditional laboratory experimentation. The exponentially increasing sea of nontraditional data (e.g., brain imaging, social media, crowdsourced norms, etc.) requires new analytic techniques for knowledge discovery. In addition, considering recent funding trends we can expect to see neuroimaging data grow at a similar rate to the growth of genetic data when we were trying to crack the human genome. SCiP is well positioned to lead the field with emerging techniques for creatively interrogating large-scale data to discover principles of behavior, and to do a better job of shepherding the field of neurotechnology—after all, brain and behavior are integrated systems.

With this in mind, I have selected a set of leading researchers in big data mining to discuss how the field is changing in my Presidential Symposium, and our Keynote Speaker this year is Chris Eliasmith, author of *How to Build a Brain*. In addition, we have a tutorial session on using the Neural Engineering Framework (Nengo), and a symposium on Massive Open Online Courses (MOOCs). All this complements an outstanding set of talk sessions and posters that really paints a bright picture of SCiP’s future. I look forward to meeting old friends and making new colleagues. If you are new to the conference, please do not hesitate to ask me or members of the Steering Committee questions or to provide your thoughts on the conference. Finally, all presenters, whether of spoken papers or posters, are invited to submit manuscripts based on your presentations to the Special Issue of *Behavior Research Methods* (see <http://scip.ws/> for details).

Enjoy SCiP 2013 and your visit to Toronto!

–Mike Jones

Table of Contents

Welcome Message.....	2
General Information.....	3
Conference Maps.....	5
Conference Schedule.....	6
Poster Presentation List.....	8
Keynote Talk.....	10
Presidential Symposium.....	11
Session A1.....	14
Session B1.....	17
Symposium.....	20
Session A2.....	23
Session B2.....	26
Nengo Tutorial.....	13
Session A3.....	30
Session B3.....	32
Poster Session.....	36

General Information

About SCiP

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

SCiP Officers

President: Michael Jones, *Indiana University*

President Elect: Kay Livesay, *Linfield College*

Past President: Ping Li, *Pennsylvania State University*

Secretary-Treasurer: Rick Dale, *University of California at Merced*

Steering Committee

Zhiqiang (Carl) Cai, *University of Memphis*

Peter Dixon, *University of Alberta*

Xiangen Hu, *University of Memphis*

Chris Koch, *George Fox University*

Randy Jamieson, *University of Manitoba*

Ulf-Diedrich Reips, *Universidad de Deusto*

Monica Riordan, *Chatham University*

Yana Weinstein, *Washington University, St. Louis*

Xiaowei Zhao, *Emmanuel College*

SCiP Time

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

Presentation Guidelines

Talks

For oral papers, presentation time will be limited to a total of 15 minutes, which includes five minutes for discussion. Sessions will be strictly timed. Presentations will be shown using Microsoft Powerpoint or Acrobat Reader on a Windows XP computer.

Posters

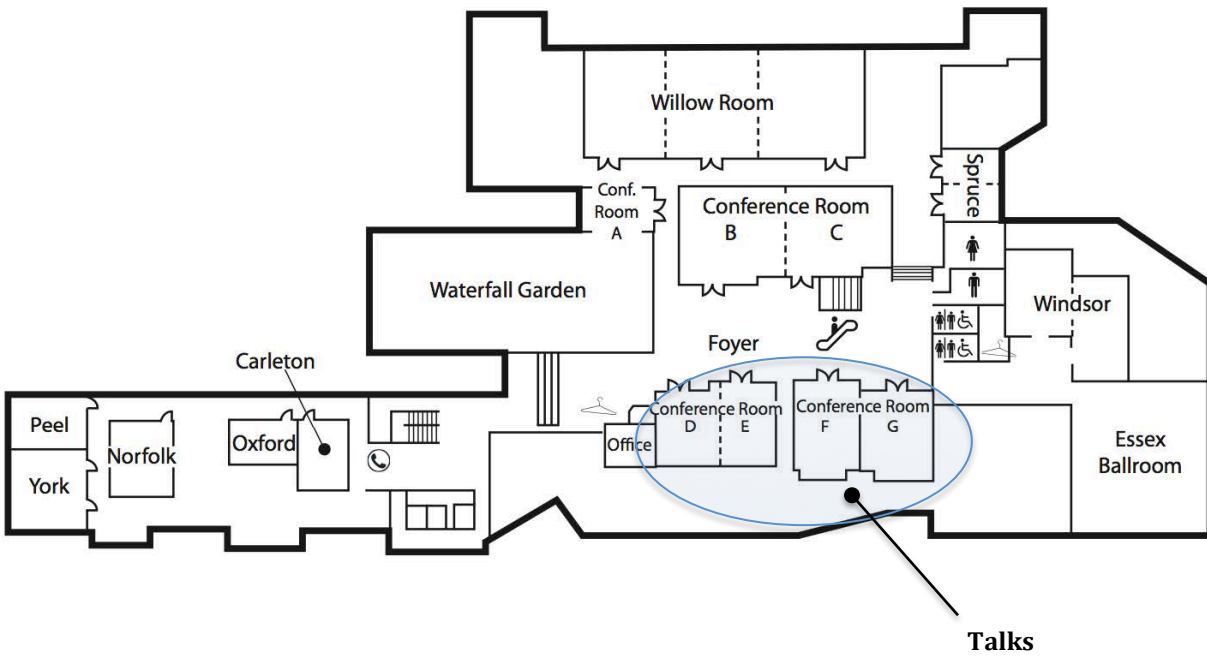
Poster presentations have the advantage of longer discussion time, less formality, and closer audience contact. The poster session will be held on Thursday, November 14th, from 1:00-2:30pm.

Keynote

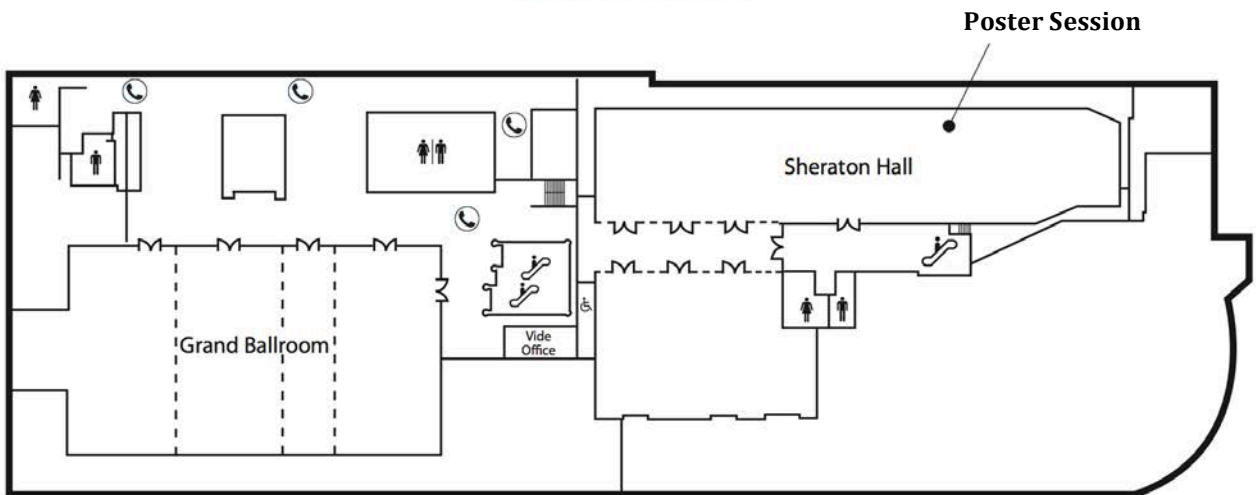
Dr. Eliasmith, *University of Waterloo*
"How to Build a Brain"

Sheraton Centre Toronto

Mezzanine



Lower Concourse



Time	Sheraton G	Sheraton F	Sheraton D & E
7:30-8:15am	Registration: Foyer		
8:15-9:30am	<p>Corpus and Discourse Analysis</p> <p>Chair: Chris Westbury</p> <p>8:15: Maskit, Bucci, & Murphy</p> <p>8:30: Zhao & Mauer</p> <p>8:45: Li, Graesser, & Cai</p> <p>9:00: Westbury & Wurm</p> <p>9:15: Paxton & Dale</p>	<p>Cognitive Neuroscience Applications</p> <p>Chair: Tim Rubin</p> <p>8:15: Rubin, Molnick, Wilcox-O'Hearn, Yarkoni, & Jones</p> <p>8:30: da Silva & McBeath</p> <p>8:45: McRae & Rabovsky</p> <p>9:00: Causeur, Lee, Cheng, & Sheu</p> <p>9:15: Bancroft, Hockley, & Servos</p>	<p>Symposium</p> <p>Educational Technologies That Support Deep Learning of Psychology in a Massive Open Online Course (i.e., MOOC) Environment</p> <p>Participants: Joordens, Cheng, Le, Lawson, Pare</p>
9:30-9:45am	Coffee Break: Foyer		
9:45-11am	<p>Educational Applications</p> <p>Chair: Carl Cai</p> <p>9:45: Cai, Hu, & Graesser</p> <p>10:00: Britt, Hughes, Hastings, Blaum, Higgs, & Wallace</p> <p>10:15: Varner, Snow, Crossley, & McNamara</p> <p>10:30: Kievit-Kylar, Jones, Grimaldi, & Karpicke</p> <p>10:45: Widmer & Wolfe</p>	<p>Web Research and Social Media</p> <p>Chair: Jon Willits</p> <p>9:45: Nye & Hu</p> <p>10:00: Reips & Garaizar</p> <p>10:15: Willits, Kievit-Kylar, Rubin, & Jones</p> <p>10:30: McClelland & Reips</p> <p>10:45: Lindemann & Fischer</p>	<p>Tutorial</p> <p>Nengo: A software suite for creating large-scale biologically realistic cognitive models</p> <p>Terry Stewart & Chris Eliasmith</p> <p>9:45am – noon</p>

Time	Sheraton G	Sheraton F	Sheraton D & E
11am-noon	Emerging Tools and Methodology Chair: Joe Houpt 11:00: Houpt, Blaha, & Burns 11:15: Blaha, Derhammer, & Dommett 11:30: Snow, Varner, Likens, Jackson, & McNamara 11:45: Garaizar, Telleria, & Reips	Semantic Modeling Chair: Cyrus Shaoul 11:00: Huang, Liao, Xie, Wang, Nye, & Hu 11:15: Fang, Malt, Ameen, & Li 11:30: Shaoul, Millin, Hendrix, Ramscar, & Baayen 11:45: Xie, Liao, Huang, Wu, Wang, Nye, & Hu	Nengo Tutorial (continued)
12-1pm	Lunch and Poster Setup		
1-2:30pm	Poster Session: Sheraton Hall		
2:30-3:45pm	Presidential Symposium: Big Data and the Future of Theorizing in Psychological Science (Sheraton F) Chair: Mike Jones Speakers: Dave Balota, Tal Yarkoni, Michael Ramscar		
3:45-4pm	Coffee Break: Foyer		
4-5pm	Keynote Address: "How to Build a Brain" (Sheraton F) Chris Eliasmith, <i>University of Waterloo</i>		
5-5:30-pm	Business Meeting (Sheraton F)		

Poster Session: Sheraton Hall; 1:00-2:30pm

Poster Presenters: Please follow the number designation for your poster setup. Poster must be set up between 12:00-1:00pm and must be removed by 3:00pm. Presenters of odd-numbered posters are to be at their posters for the first 45 minutes, and presenters of even-numbered posters at theirs for the second 45 minutes of the poster session.

1. Li, Clewley, Walker, Cai, & Graesser: Speech Act in Epistemic Game: In and out of the Context
2. Chan, Magliano, & Armstrong: Do Students in Developmental Reading Literacy Programs Process Text Differently than Traditionally Admitted Students?
3. Li, Graesser, & Cai: Evaluation of Google translation at multi-textual level
4. Murphy, Maskit, & Bucci: Measuring the trait component of speech disfluency using computerized language measures
5. Mauer & Zhao: Examining noun-bias in Spanish using the CHILDES database
6. Liao & Wang: Exploring the role of stimulus-response compatibility on the spatial Stroop effect
7. Thanasuan & Mueller: Crossword Expertise as Recognition Making: An Artificial Intelligence Approach
8. Cha, Chen, Chang, & Sung: The Development of An Automated Analyzer for Chinese Text
9. DeSoto: Collecting confidence ratings in cognitive psychology experiments: Effects of scale and entry method
10. Koch: Using SuperLab to Improve Understanding of Research Design
11. Kievit-Kylar, Jones, Todd, & Ahn: Typicality and Age of Acquisition across Sensory Domains
12. Birnbaum: Website with Helpful Resources for Learning How to Conduct Research via the Internet
13. Rubin & Jones: Accounting for the nature of human response selection in evaluating semantic models
14. Lan: Study of the communication strategies used by Australian CFL learners in Second Life
15. Hays & Spritzer: REcycled Sas PrEdiCTions (RESPECT)
16. Maples & Burgess: Emoticons Behaving Ambiguously
17. Burgess & Cao: The Use and Connotation of the N-word in Twain's Novels and in Contemporary Society

18. Carroll, Mendoza, Kozich, Estep, & Burgess: The Validity of Osgood's Emotional Word Features
19. Weil, Fisher, & Wolfe: Boring Begets Ignoring: The Emotions of Learning with an Intelligent Tutoring System
20. Dolgov, Nearents, Graves, Volkman, & Schwark: A Wii bit of help: Cooperative gaming increases number of subsequent spontaneous helping between the players
21. Cedilos & Wolfe: The Use of Explanation and Argumentation in Intelligent Tutoring System Dialogues
22. Huette, Anderson, & Spivey: Probabilistic negation: computer mouse-tracking reveals non-linear distortions of truth perception when negated
23. Shubeck, Germany-Schubeck, Craig, Dev, Hu, Koch, & Heinrichs: VCAEST: Training Facilitated by an ITS embedded in a Virtual World
24. Matsuka: Using multiobjective optimization techniques to effectively explore statistical models

How to Build a Brain

SCiP 2013 Keynote Address



Chris Eliasmith
University of Waterloo

Abstract

Recent high-profile brain simulations, including those of the \$1 billion Euro Human Brain Project, are very large and complex. However, they do not exhibit interesting behaviours and so are difficult to compare to much of what we know about the brain. In this talk, I describe the methods and tools used to construct what is currently the world's largest **functional** brain simulation, called the Semantic Pointer Architecture Unified Network (Spaun). I demonstrate the variety of behaviors the model exhibits and show that it is similar in many respects to human and animal behaviour. I argue that constructing such large-scale simulations is critical for advancing our understanding of neural and cognitive function.

Big Data and the Future of Theorizing in Psychological Science

SCiP Presidential Symposium



Introduction

Michael N. Jones (Chair)
Indiana University

While laboratory research is still the backbone of tracking causation among behavioral variables, more and more research is now letting experimental control go in favor of large-scale and real-world data. Online crowdsourcing, mega databases, social media, and genetic/neural data are allowing researchers to evaluate theoretical models and make new discoveries at a resolution not previously possible. Knowledge gleaned from large-scale behavioral data is dependent on developments in analytic techniques. Although the causal links among variables are murkier, they are still possible, and the scale of data allows connections among neural, linguistic, and behavioral data to paint a more complete and realistic picture of psychological mechanisms. This symposium features leading researchers from different areas of psychology all using big data to address theoretical questions in psychology. We hope to have a fruitful discussion of the role of big data in psychology's future, SCiP's role in developing analytical techniques to glean knowledge from these data, and some of the pitfalls that must be avoided.



Megastudies in word processing: What have we learned?

David Balota
Washington University in St. Louis

The vast majority of research in visual word processing has involved standard factorial studies in which investigators cross targeted variables (e.g., high vs low-frequency words and long vs short words). These studies have provided a rich literature to build models of word recognition. In contrast to the factorial approach, there has been an increasing utilization of large databases (megastudies) to inform models of visual word recognition. This talk will review the empirical and theoretical contributions from the megastudy approach and discuss the advantages and potential limitations of this approach. Attention will be given to the user interface for these databases, and more recent powerful methods to develop such databases.



An open framework for large-scale, automated synthesis of the human functional neuroimaging literature

Tal Yarkoni
University of Texas at Austin

The explosive growth of the human neuroimaging literature has led to major advances in understanding of human brain function, but has also made aggregation and synthesis of neuroimaging findings increasingly difficult. To address this problem, we recently introduced a highly automated brain mapping framework called Neurosynth that uses text mining, meta-analysis and machine learning techniques to automatically synthesize much of the functional neuroimaging literature. In this talk I review a program of research that uses this framework to address a range of theoretical and methodological problems in cognitive neuroscience. I review several different applications, focusing particular attention on how a federated and open infrastructure for functional neuroimaging can benefit the fields of psychology and neuroscience.



Psychological theory: Back to the future with big data

Michael Ramscar
Tübingen University

The "cognitive revolution" brought together a number of disciplines with very different intellectual approaches. Whereas linguistics always tended towards a combinatoric approach in its theories, historically, psychologists were much preoccupied with discriminative learning, and the role it appeared to play in shaping apparently combinatoric cognitive processes. The discriminative perspective fell out of favour in the latter part of the last century, and is barely even considered when contemporary psychologists theorize about high-level cognition. In this brief talk, I'll try to show why I think big data can reinvigorate discriminative theorizing, and in turn, the scientific study of mind, simply because many seemingly intractable psychological problems look much less mysterious when viewed from a discriminative rather than combinatoric perspective. I'll illustrate this point with examples from recent work examining name memory and cognitive development across the lifespan.

Nengo: A software suite for creating large-scale biologically realistic cognitive models

SCiP Tutorial Session: 9:45am – noon, Sheraton D & E



Terry Stewart
University of Waterloo



Chris Eliasmith
University of Waterloo

We recently developed Spaun, the world's first simulated brain that is capable of performing multiple cognitive tasks, including serial working memory, reinforcement learning in a bandit task, and pattern completion in Raven's Matrices. To achieve this using 2.5 million simulated spiking neurons we have developed a theoretical framework and a software toolkit that allows us to take high-level theories of the basic processes of cognition and determine how neurons may be organized and connected to result in the desired behaviour. In this tutorial, we introduce you to the Neural Engineering Framework and the Semantic Pointer Architecture, our theoretical tools that support this effort. At the same time, we will lead a hands-on demonstration of Nengo, our open-source software that creates and runs these neural simulations <<http://nengo.ca>>. The goal is to allow attendees to use these tools to develop their own software simulations of cognitive systems for their own research.

8:15-9:30am Parallel Session 1

Session A1: Corpus and Discourse Analysis (Sheraton G)

Chair: Chris Westbury, University of Alberta

8:15 A new text analysis system and its application to episodic memory

Bernard Maskit (daap@optonline.net), Stony Brook University
Wilma Bucci (wbucci@optonline.net), Derner Institute, Adelphi University
Sean Murphy (smurphy1@gmail.com), New York Psychoanalytic Society and Institute

The Discourse Attributes Analysis Program (DAAP) compares lexical items in a text with weighted or unweighted dictionaries. Using a local weighted averaging procedure, DAAP produces smooth curves to model fluctuations of psychological variables, across either time or word count. DAAP then produces new measures of within text variation in dictionary matches and covariations between pairs of dictionaries using these curves. Here we apply the Weighted Referential Activity Dictionary (WRAD) (Bucci & Maskit, 2006), developed to measure degree of connection between non-verbal experience and language. Our new measure here is HWP, the proportion of words for which the smooth WRAD curve lies above its neutral value. Variation in WRAD is associated with degree of detailed imagery and narrative; we expect HWP will be associated with ratings of episodic memory. Using data provided by Schacter (Addis, Wong & Schacter, 2008), consisting of 16 narratives each of past or future events from 32 subjects, we compared HWP with a measure of episodic memory developed by Levine et al. (2002). The transcribed responses were separated into "details", scored as either "internal" (explicit details about the main event being described), or "external" (other details). The proportion of internal to total details compared with HWP, averaging across all 16 prompts, yielded a high Pearson correlation (N=32, $r=.705$, $p<.001$). The 26 subsets yielded 20 significant correlations ranging between $r=.377$ and $r=.860$, providing evidence for HWP as a computerized measure of episodic memory. Differences between concepts underlying referential activity and episodic memory are discussed.

8:30 Investigating "Noun bias" across three languages: A computational study based on parallel corpora

Xiaowei Zhao (zhaow@emmanuel.edu), Emmanuel College
Meaghan Mauer (mauerm@emmanuel.edu), Emmanuel College

In this study, we present a neural network model called DevLex-II (Li, Zhao & MacWhinney, 2007; Zhao & Li, 2009) to simulate the early vocabulary development in three different languages: English, Spanish and Chinese. We focus on how the different lexical composition patterns in the three languages can emerge, develop and change when the learner acquires an increasing number of words. In particular, our analysis focus on the so-called “Noun-bias”, a hypothesized universal pattern that shows a predominance in the number of nouns, as compared with other categories of words, in the child’s early lexical composition. In order to conduct a more accurate comparison across the three languages, we used the translations of the same material as the basic training set of our analysis (i.e. parallel corpora in linguistics). Specifically, we used the book of Grimm's Fairy Tales, and great care was taken to ensure that the exact translations of the book were used for our modeling simulation. The simulation results are further compared with those extracted from empirical studies, such as lexical norms and children’s speech.

8:45 **A Comparative Study on Measures of Text Formality**

Haiying Li (hli5@memphis.edu), University of Memphis
Arthur Graesser (graesser@memphis.edu), University of Memphis
Zhiqiang Cai (zcaic@memphis.edu), University of Memphis

Formality has long been of interest in the study of discourse, with periodic discussions of the best measure of formality and the relationship between formality and text categories. In this research, we explored what features predict formality as humans perceive the constructs. The corpus consisted of 1158 discourse samples published in the Collected/Selected Works of Mao Zedong, which were classified into the following categories: conversations, speeches, letters, comments, published articles, telegrams and official documents. Two formality models were constructed to measure formality: (1) 5 out of 7 factors extracted from 73 linguistic, discourse and psychological features through a principle components analysis and (2) 5 out of 8 word classes. Comparisons were made between component measures, word class measures, as well as other measures, when predicting human formality rating. Findings showed weighted component measure best predicted human perceptions of formality than previous formality metrics, as it provided an objective, unbiased, and panoramic measure of formality. Moreover, this paper examined the predictability of formality on variation in text categories. Findings revealed human formality judgments could considerably predict the variation in text categories. Furthermore, the comparisons of different formality metrics in term of text categories demonstrated that weighted component measure of formality could better and stably explain variations in text categories in generalization than other formality metrics, especially the distinction of well-prepared texts from impromptu texts. This study is the first to collect human judgments of formality for lengthier stretches of texts and to systematically assess alternative automated metrics of formality.

9:00 **What's the Most Dangerous, Least Useful Thing?**

Chris Westbury (chrisw@ualberta.ca), University of Alberta
Lee Wurm (lee.wurm@wayne.edu), Wayne State University

The dimensions of danger and usefulness have been shown to play a role in lexical access (see review in Wurm, 2007). The extent to which a word loads on either dimension has been measured with human judgments, masking the underlying motivation of the decisions. The dimensions have been assumed to reflect approach/avoidance emotional evaluations. We introduce an empirically grounded stochastic computational method for estimating danger and usefulness using co-occurrence distances obtained from a large text corpus using the open source model HiDEx (Shaoul & Westbury, 2010). We analyzed 100,000 unique regression models to predict each measure, with each composed of a frequency measure and computationally-derived co-occurrence distances from eight emotional labels selected at random from a pool of 77 terms that have been proposed in the published emotion literature as 'basic emotions'. By analyzing the 77 terms as a function of their predictive power across this large set of random models, we are able to quantitatively motivate a small set for predicting the two dimensions of danger and usefulness. We demonstrate that the predictors identified in this way are strongly predictive of human judgments, present qualitative analyses suggesting that they generalize to a large set of over 36,000 words, and discuss the statistical distribution of the measures across that large set.

9:15 **Keeping Time: Dynamics in Text Analysis**

Alexandra Paxton (paxton.alexandra@gmail.com), University of California, Merced
Rick Dale, University of California, Merced

From clinical transcripts to first-person narratives, text analysis plays a central role in a wide variety of research areas. These sources often incorporate a temporal element, sometimes with multiple possible time scales: For instance, a clinician may be interested in analyzing her client's language use for affective changes over several months or within a single session. Temporal effects can be washed out by aggregating across time scales, preventing researchers from seeing intriguing patterns within their data that occur at faster time scales. To that end, we here present the by-word longform (BWLf) or "B(eo)W(u)LF," a flexible data format intended to facilitate multi-level language analysis, and demonstrate its potential uses with real-world data.

Session B1: Applications to Cognitive Neuroscience (Sheraton F)

Chair: Tim Rubin, Indiana University

8:15 Neurotopics: A toolbox for learning the associations between linguistic features and regions of brain activation from a corpus of neuroimaging publications

Timothy Rubin (timrubin@indiana.edu), Indiana University
Jessica Molnick, University of Colorado, Boulder
Amber Wilcox-O'Hearn, University of Colorado, Boulder
Tal Yarkoni, University of Texas at Austin
Michael Jones, Indiana University

As more and more neuroimaging data are published, methods for performing large-scale analyses of existing publications have become increasingly important for (1) summarizing existing knowledge of the functional roles of different brain regions, and (2) suggesting new functional relationships that may not be observable in individual experiments. Here, we present a method for automatically learning associations between linguistic features and regions of brain activation using a large corpus of neuroimaging publications. This method employs a novel application of conditional Latent Dirichlet Allocation, which learns a probability distribution over linguistic features for each brain region within any given functional/anatomical map of the brain. These probability distributions can be useful for summarizing the functional relationships between neural activation and behaviors in a corpus, as well as for suggesting novel functional roles for brain regions. These modeling techniques are implemented in Neurotopics--an open-source toolbox that is currently under development. We present exploratory results using this model, including a comparison of the linguistic features the model associates with each brain region when different mappings of brain regions are employed. We additionally present an extension of the model that uses linguistic features to find novel subdivisions of brain regions in an unsupervised manner, and demonstrate that this can recover spatially separated subdivisions known to be functionally distinct.

8:30 Generating machine commands from EEG signals

Flavio da Silva (flavio.dasilva@asu.edu), Arizona State University
Michael McBeath (michael.mcbeath@asu.edu), Arizona State University

We used an Emotiv EPOC® electroencephalographic brain-computer interface (EEG-BCI), and trained 24 participants to control movement of a cursor on a computer screen directly from EEG signals. We explore two methods of generating EEG signals and algorithms to transform them into machine commands. Two methods of mental imagery were used to generate EEG control signals. (1) Embodied Imagery: Twelve participants imagined

using their arms and fingers to move the computer cursor. (2) Abstract Imagery: Twelve participants imagined abstract, disembodied concepts (justice, infinity, etc.). We used a simple feed-forward artificial neural network (ANN) to map EEG signals into machine commands. Network inputs consisted of 14 spectrograms generated using short-time interval fast Fourier transforms (SFFTs) on EEG recordings over time (one spectrogram for each of the 14 electrodes). We trained the ANN using simple back-propagation error-correction. Overall, our method using spectrograms as inputs to a simple ANN to transform EEG signals into machine commands compares favorably with the proprietary mapping used by Emotive, Inc. However, individual performance varied considerably, exhibiting a negative correlation between those performing well using our method and those using Emotive's proprietary method. The same inverse relationship held true independent of whether individuals used embodied or abstract imagery to generate EEG signals. The varying patterns of performance suggest our method of mapping EEG signals into machine commands was fundamentally different from Emotive's method. In summary, results support that for a given population, developing multiple methods of mapping EEG signals into machine commands can improve overall performance of brain-computer interfaces.

8:45 **Simulating the N400 ERP component as semantic network error: Insights from a feature-based connectionist attractor model of word meaning**

Ken McRae (kenm@uwo.ca), University of Western Ontario
Milena Rabovsky, University of Western Ontario

The N400 ERP component is widely used in research on language and semantic memory. Although the component's relation to semantic processing is well-established, the specific mechanisms underlying N400 generation are currently unclear. We explored the mechanisms underlying the N400 by examining how a connectionist model's performance measures covary with N400 amplitudes. We simulated seven N400 effects obtained in human empirical research. Network error was consistently in the same direction as N400 amplitudes, namely larger for low frequency words, larger for words with many features, larger for words with many orthographic neighbors, and smaller for semantically related target words as well as repeated words. Furthermore, the repetition-induced decrease was stronger for low frequency words, and for words with many semantic features. In contrast, semantic activation corresponded less well with the N400. Our results suggest an interesting relation between N400 amplitudes and semantic network error. In psychological terms, error values in connectionist models have been conceptualized as implicit prediction error, and we interpret our results as support for the idea that N400 amplitudes reflect implicit prediction error in semantic memory.

9:00 **Significance testing of massive correlated correlations in ERP data analysis**

David Causeur (causeur@agrocampus-ouest.fr), Applied Mathematic Department, Agrocampus Ouest
Yuh-shiow Lee (psyysl@ccu.edu.tw), National Chung Cheng University
Kuan-Hao Cheng (kuanhao@gmail.com), National Chung Cheng University
Ching-Fan Sheu (csheu@mail.ncku.edu.tw), National Cheng Kung University

Many event-related potentials (ERP) studies are motivated by a desire to relate the time courses of mental events under investigation to some relevant behavioral measures. Unfortunately, testing for associations between ERP and behavior suffers from low statistical power when p-values of correlations are corrected for multiple comparisons. One common practice to circumvent the multiplicity problem is to select a window size before averaging ERPs across successive time-windows for a particular electrode of interest. However, simulation studies (e.g., Kilner, 2013) have shown that post-hoc data aggregating strategies tend to bias toward rejecting the null hypothesis resulting in inflating the number of false positives, particularly when temporal correlation is present, which is the case for real ERPs. In this work we extend the recently developed dynamic factor-adjusted multiple testing procedure under dependence (Causeur, Chu, Hsieh, & Sheu, 2012) to massive testing of correlated correlations between ERPs and behavioral measures. For comparison, time points at which ERPs are significantly associated with behavior are detected by the false discovery rate control (Benjamini and Hochberg, 1995) of the factor-adjusted p-values. We demonstrate the advantages of this novel approach by stimulation and by analyzing associations between recognition memory performance (hit rate minus false alarm rate) and ERPs collected from a study of directed forgetting using a (modified) stop-signal paradigm.

9:15 **Scalar short-term memory: A paradigm for integrating computational models with psychological and neural data**

Tyler Bancroft (banc6110@mylaurier.ca), Wilfrid Laurier University
William Hockley (whockley@wlu.ca), Wilfrid Laurier University
Philip Servos (pservos@wlu.ca), Wilfrid Laurier University

Scalar short-term memory is memory for stimuli where the salient stimulus property is scalar or unidimensional. Examples include tactile memory for vibrational frequency, auditory memory for pure tones, and visual memory for the flicker rate of LEDs. These tasks are easily implemented in human research, and have proven useful for testing theories of interference in short-term memory, but have also been extensively studied at a neural level in primates. The rich literature on the neural correlates of scalar memory has

allowed the construction of computational models of the neural systems underlying scalar memory, and the application of those models to human performance. We have recently demonstrated that while such models are based on data obtained from non-human primates, they are also capable of replicating data from human behavioural experiments. We will review recent developments in the study of scalar memory, both computational and experimental, and argue that scalar short-term memory is an ideal model system for integrating data from human and non-human psychology and neuroscience, and for the development and application of computational models to that data.

Session C (Symposium): Educational Technologies That Support Deep Learning of Psychology in a Massive Open Online Course (i.e., MOOC) Environment (Sheraton D & E)

Chair: Steve Joordens, University of Toronto at Scarborough

Overview: In June and July 2013, Professor Steve Joordens offered an 8-week Introduction to Psychology course on the Coursera platform, sponsored by the Bill and Melinda Gates Foundation. Over 45,000 students were active in the course at some point, and over 11,000 remained active to the end. Rather than using the assessment tools native to the Coursera platform, Professor Joordens and his students utilized tools they have been designing, tools specifically created to support deep learning (i.e., critical thought, creative thought, reflective thought, communication, hands-on active learning, etc.).

In this symposium, we will provide an overview of the course, the tools we implemented and the resulting findings. Specifically, Professor Joordens will first provide an overview of the course, including demographic statistics, structure, and the timing of the assessments and activities. Across four subsequent presentations his students will then highlight specific assessments or issues of interest, will describe them and their pedagogical basis in detail, and then will present research findings obtained within the MOOC context.

8:15 Introduction to Psychology as a Massive Open Online Course

Steve Joordens, and the entire UTSC MOOC team

There are many challenges associated with offering Introduction to Psychology as an 8 week MOOC. In terms of content, one must decide how to take what normally would be 48 or more hours of lecturing, and repackage it as 64 videos that are roughly 10 to 15 minutes each (i.e., 16 hours overall) while still covering all the critical content areas. In addition, it is then necessary to consider the other aspects of the course itself including assessments and activities. Finally, while the MOOC is ongoing, one must decide how it will be supported, and what measures will be taken to create and sustain a true learning community on such a large and varied scale. In

my presentation I will describe the MOOC we created in general terms, and I'll provide some descriptive statistics ranging from demographics of the student body, to retention and general reaction of the students. I will also highlight the instructor experience and the workload involved.

8:30 Assessing and Building Knowledge Using mTuner

Cho Kin Cheng and Rob Walker

The primary goal of MOOCs is to make quality education freely available to anyone who is interested. Thus learning is primarily driven by intrinsic motivations, and the goal of the course providers is to create a course that maximizes learning. However, students who successfully demonstrate their mastery of course content can also earn a Certificate of Accomplishment necessitating some form of assessment. The Coursera platform includes a tool for generating multiple-choice tests but we chose to use a tool we developed, a tool that harnesses all current research on assessment for learning to not only assess knowledge, but also to strengthen learning by cementing accurate knowledge and correcting misconceptions, in an experiential context. This tool, called mTuner, works as follows. First, only the question is displayed and students are asked to free recall the correct answer. They are then shown the alternatives and given a specified amount of time to choose one. If they are correct, they are immediately congratulated. If they are incorrect they are returned to a portion of the relevant lesson, allowed to watch it for a minute, and are then returned to the alternatives for a second chance for half marks. If they still get it wrong then they are shown the correct answer. In all cases the question ends by presenting an additional statement that explains why the correct answer is correct. In our presentation we will briefly demonstrate mTuner, and then will present both qualitative and quantitative data related to its use in our MOOC course.

8:45 Exercising Critical Thought, Effective Communication and the Formation of Connections Using peerScholar

Dwayne E. Pare

PeerScholar is a tool developed within our lab to allow students to gain effective practice in deep thought (i.e., critical thinking, creative thinking and reflective thinking) and efficient, effective communication in any sized learning environment. It harnesses the research-supported power of peer- and self-assessment to allow students to help one another, and to learn a great deal in the process. This tool is now being used in Universities and Schools the world over, but our course was the first time it was used within a MOOC context thereby representing the largest scale implementation. In our presentation we briefly describe and demonstrate peerScholar, highlighting

the pedagogy underlying its use, and the general notion of the value of “exercising while assessing”. The specifics of the implementation used in this course are then presented to show how a MOOC implementation compares to what is possible when TAs are available to do some grading. Finally, we present both qualitative and quantitative research findings relating to both the efficacy of the tool and students’ reactions to it.

9:00 **Understanding Science by Testing Theories using Digital Labcoat**

Ainsley Lawson

There is a strong research base supporting the efficacy of active learning strategies; activities that require students to use the knowledge they have learned shortly after learning it. Digital Labcoat was created for this purpose, specifically to deepen the learning of the scientific method, including the role played by statistics in terms of answering theoretical questions. An optional Digital Labcoat activity was made available to students in our MOOC course just after three lectures devoted to the scientific method. In this presentation we will demonstrate the four stages of Digital Labcoat; filling out a questionnaire, mining subsets of the questionnaire data in search of interesting contrasts or relationships, attempting to replicate the interesting findings of others, and finally providing and voting on theoretical accounts of the most interesting replicable findings. Within this context we will then present both qualitative and quantitative data reflect the impact of Digital Labcoat on learning.

9:15 **Who Succeeds, and What Do We Really Mean by Success?**

Ada Le

Massive open online courses (MOOC) are free-of-charge online courses offered by top universities around the world, and are typically aimed at a large-scale audience, which include students of all ages, education level, and socio-economic background. Moreover, MOOC is a method for delivering high-quality educational content without any limits or restrictions on attendance. As such, activity level and retention rate for these courses tend to exhibit a rapid decline (e.g. down to 3 – 4%) as the course progresses. Nevertheless, full attendance and satisfactory performance are typically rewarded with a certificate at the end of the course. Therefore, one pressing question arises: What causes students to either drop out or continue in the course, and if they continue, what predicts their activity and performance level? Here we examine this question by having collected data from over 12,000 students in an Introduction to Psychology MOOC at the University of Toronto. This data included (1) measures of students’ personality and motivation characteristics; along with their (2) course grades; (3) activity level in watching lecture videos and posting on the discussion forum; and (4)

date of registration and last access, reflecting retention duration in the course. We submitted these measures to a multiple regression analysis, and found that student personality and motivation characteristics predicted the students' grades, activity level, and retention duration in the course. This research will be paramount for future efforts aimed at improving the activity level and retention rate in MOOCs, and at improving student satisfaction in online courses in general.

9:45-11:00am Parallel Session 2

Session A2: Educational Applications (Sheraton G)

Chair: Carl Cai, *University of Memphis*

9:45 ASAT: AutoTutor Script Authoring Tool

Zhiqiang Cai (zca@memphis.edu), University of Memphis
Xiangen Hu (xhu@memphis.edu), University of Memphis
Arthur Graesser (agraesser@memphis.edu), University of Memphis

The agent builder is used to specify agents involved in conversations. The information about an agent includes names, gender, age and title. In addition to that, a script author can create customized canned expressions such as greeting, positive feedback, neutral feedback, negative feedback, etc. The canned expressions can help personalize agents. Script editor is used to create materials for a conversation, including speeches for each agent and expected responses from human users. The rule design tool allows a conversation designer to create various conversation patterns that are used to perform specific tasks. The conversation engine is a web service program that is used in ASAT for testing scripts. The engine can be integrated into tutoring and game environments. ASAT testing interface with web based avatars is used to valid scripts.

10:00 Evaluating Argumentative Essays for Causal Explanations about Complex Scientific Phenomena

M. Anne Britt (britt@niu.edu), Northern Illinois University
Simon Hughes (simonhughes22@hotmail.com), DePaul University
Peter Hastings (peterwh@cs.depaul.edu), DePaul University
Dylan Blaum (dylblaum@gmail.com), Northern Illinois University
Karyn Higgs (karynhiggs@yahoo.com), Northern Illinois University
Patty Wallace (pwallace@niu.edu), Northern Illinois University

Recent science and literacy standards require students to use multiple sources of information to understand explanations for phenomena and to

use data to support their own explanations. Thus, there is a critical need for a method of evaluating students' explanations and argumentative support based on scientifically-important criteria (e.g., coherence, completeness, accuracy). In a recent study, 250 middle and high-school students wrote an essay explaining the causes of global warming using information from 7 documents. Humans evaluated the essays for the extent to which each purported cause (e.g., increased burning of fossil fuels) was explicitly connected to the target effect (increase in global temperatures) and the number of intervening factors mentioned (e.g., increased CO2 trapping heat). Support statements (e.g., rate of current burning) were also counted. Inter-rater reliability was high ($Kappa = 0.85$) and the method was useful in discriminating essays based on coherence and completeness. In fact, 38% of the concepts and 66% of the support statements were not explicitly connected. Thus, a simple scoring method of identifying concepts without analyzing their structural relationship would not have been useful for evaluating these criteria. Then, we compared human scoring with automatic methods. A support vector machine classifier and a nearest-neighbor classifier were effective in detecting the causes and effects within a sentence. We are currently evaluating automated techniques to capture cause-effect relationships within and across sentences. In this talk, we will describe in detail both the human scoring and the automatic techniques and discuss the strengths and weaknesses of each technique.

10:15 **Whad'ya say? Examining Linguistic Correlates of Students' Abilities**

Laura Varner (laura.varner@asu.edu), Arizona State University
Erica Snow (Erica.L.Snow@asu.edu), Arizona State University
Scott Crossley (sacrossley@gmail.com), Georgia State University
Danielle McNamara (dsmcnamara1@gmail.com), Arizona State University

National assessments demonstrate that many students write below proficiency. To remediate this problem, researchers have investigated cognitive skills and knowledge sources that contribute to variations in students' writing proficiency (e.g., Kellogg, 2001). Despite some success identifying these individual differences, it remains unclear how these factors manifest themselves in terms of the properties of students' writing. We examined the relations between students' writing characteristics and cognitive abilities at a more fine-grain size. Specifically, we utilized natural language processing techniques to determine whether linguistic indices of students' essays were predictive of their scores on a variety of cognitive assessments. University students ($n = 108$) wrote timed, prompt-based essays in response to an SAT-style prompt. Additionally, they completed a battery of measures (e.g., working memory span, text inferencing ability) aimed at assessing their lower- and higher-level cognitive skills, and their vocabulary knowledge. Coh-Metrix (McNamara et al., in press), an automated text analysis tool, was then used to calculate the linguistic

properties of the students' essays. Correlation and regression analyses were calculated between scores on the cognitive tests and the linguistic properties of the essays. Results indicated that automated linguistic indices significantly predicted variance in scores on the cognitive tests. In particular, vocabulary knowledge was significantly related to the sophistication and specificity of students' word choices. Conversely, higher-level cognitive processes were more strongly related to essay cohesion. Overall, results of this analysis confirm that automated text tools can detect and provide important information about the role of cognitive abilities in students' writing ability.

10:30 **A Semantic Model Supported Tutoring System for Retrieval Based Learning**

Brent Kievit-Kylar (bkievitk@indiana.edu), Indiana University
Michael Jones, Indiana University
Phil Grimaldi, Purdue University
Jeff Karpicke, Purdue University

In this paper we discuss a series of augmentations to QuickScore (PJ Grimaldi & Karpicke, 2013) to enable this automated scoring system to include semantic relationships. QuickScore is an algorithm based on the principle of retrieval learning, notably that the optimal form of learning is through repeated testing and retrieval. To this end, a feasible implementation of this principle requires continuous feedback for the learner (a resource and time expensive task for a teacher). An automated scoring system can be used to augment the ability of a teacher by providing automated scoring, as well as aggregating semantic representations of an entire class to provide areas of overall difficulty and misunderstanding within the group or sub-groups. The QuickScore algorithm simply produces a hit-rate between a labeled "correct" answer and a target response by comparing each word individually and accounting for variation such as misspellings, prefixes and postfixes and superficial letter features (ie letter case). While this works well for explicit retrieval type questions, the unforgiving nature of the matches is poor at scoring any form of integrative response. To counter this, we propose integrating a semantic space model into the framework. Such an addition allows partial matching between word as well as comparisons of areas in semantic space, thus matching similar meanings instead of similar words.

10:45 **Argumentation and Intelligent Tutoring in an Ill-Structured Domain**

Colin Widmer (widmercl@miamioh.edu), Miami University
Christopher Wolfe (wolfecr@miamioh.edu), Miami University

Intelligent Tutoring Systems (ITS) have been shown to lead to large learning

gains by encouraging learners to elaborate on what they have learned through self-explanation. While learning gains from self-explanations have been studied, there is evidence that writing arguments can also lead to learning gains, and that writing arguments may be particularly appropriate to learning about ill-structured domains. In this study 123 participants received four interactive lessons on topics in the ill-structured domain of civics in AutoTutor Lite, a web-based ITS that interacts with learners in natural language to help them write reflections on each lesson. Participants were randomly assigned to one of four conditions: a didactic lesson alone with no interaction, a didactic lesson followed by an interaction to write an explanation about the topic, a didactic lesson followed by an interaction to write an argument about the topic, or a control condition with an irrelevant lesson. All participants who received the didactic content did significantly better on a knowledge test than the control, but interacting with the tutor in either manner did not increase performance. However, interacting to create arguments did have an effect on participants' ability to distinguish higher quality arguments about some topics. Specifically, participants who wrote arguments showed a greater sensitivity to the value of warrants, i.e. the linking of supporting data to the argument's claim, than those who wrote explanations. We also present data from essays written by each participant and discuss results in terms of a potential role for argumentation in ITS.

Session B2: Web-Based Research and Social Media (Sheraton F)

Chair: Jon Willits, *Indiana University*

9:45 Integrating Natural Language Tutoring into Existing Web-Based Worked Examples Using Sharable Knowledge Objects (SKO)

Benjamin Nye (benjamin.nye@gmail.com), Institute of Intelligent Systems, University of Memphis

Xiangen Hu, Institute of Intelligent Systems, University of Memphis

Worked examples have been shown to improve both procedural skills and conceptual knowledge of mathematics, such as geometry and algebra. In many cases, worked examples outperform practice problems because they draw attention to key features (reducing cognitive load), which allows deeper processing of the domain knowledge. Unfortunately, learners often do not engage in deep processing of examples, resulting in shallow understanding. To support deeper understanding from worked examples and other non-interactive learning materials, a new mixed-initiative natural language tutoring system called SKO-ITS (Sharable Knowledge Object Intelligent Tutoring System) is being developed. This paper describes the design of this intelligent tutoring system and the process for creating and

delivering tutor-supported worked examples for Algebra I. SKO-ITS is based on service-oriented design principles, where the tutoring system is actually a flexible composition of independent services (e.g., animated talking agents, semantic analysis, student model, dialog interpreter, etc.). This system is based on AutoTutor, an intelligent tutoring system that has previously demonstrated learning gains on conceptual physics, computer literacy, and other topics. SKO are applying AutoTutor-style dialogs to enhance non-interactive web content by using animated pedagogical agents to support the learner with tutoring discourse in natural language (free text). These tutoring agents can use pedagogical strategies such as deep questions to foster deeper understanding from existing algebra worked examples. Innovations on methods to initiate and monitor these dialogs will be discussed, as well as techniques to gracefully respond to unconstrained user input.

10:00 **A set of JavaScript tools for Web-based research**

Ulf-Dietrich Reips (u.reips@ikerbasque.org), University of Deusto;
IKERBASQUE, Basque Foundation for Science
Pablo Garaizar (garaizar@deusto.es), University of Deusto

During the last two decades Internet-based research was established as a viable alternative to traditional computer-based and non-computer-based research. Most of Internet-based research is conducted via a Web user-agent, i.e. a Web browser. Web browsers and their underlying technologies evolve over time. While some technologies, such as Java, Flash, and Authorware have lost their appeal and are becoming outdated, other technologies such as HTML and Javascript have survived and continue to work in all Web browsers. Hence, these technologies are most suited for Web-based research. In the current paper we present a set of JavaScript tools for use in such research. These tools work in all web browsers, they fulfill the following functions and can be used in modular ways. (1) Identification of user's system (e.g., referrer, operating system, user-agent, resolution, width and height of screen), (2) millisecond timing of page display (3) millisecond measurement of time of display (4) millisecond measurement of time taken to build web page on user screen, (5) create a random session identifier and (6) carry it over a series of web pages, (7) record mouse arrow position on page. We explain how to send/receive data to/from a server asynchronously via AJAX or Server-Sent Events, how to save a backup of the results using the FileSystem or Web SQL APIs, how to create threads to compute partial results during an experiment or survey, and how to control the number of times a research study is done from a browser using super cookies.

10:15 **Semanticore.org: A web-based portal for semantic models and psycholinguistic data**

Jon Willits (jwillits@indiana.edu), Indiana University
Brent Kievit-Kylar (bkievitk@umail.iu.edu), Indiana University
Tim Rubin (tim.rubin@gmail.com), Indiana University
Michael Jones (jonesmn@indiana.edu), Indiana University

Recent years have seen a rapid increase in the number of semantic models and big data approaches for analyzing psycholinguistic data. Until now, these techniques and the associated data have been spread across disparate labs and sources. Here, we present Semanticore.org, a web-based portal for semantic models and psycholinguistic data. Semanticore will focus on making four resources available at one location. First, Semanticore will provide source code, documentation, and a tutorial repository for popular semantic space models like LSA, HAL, BEAGLE, and Probabilistic Topic Models. Second, Semanticore will feature pre-computed semantic space data for a wide range of corpora and semantic space models. Third, Semanticore will serve as a repository for other kinds of psycholinguistic data. This will include corpus statistics like word frequencies, co-occurrence probabilities, and measures of semantic diversity. It will also include a repository for a wide range of normative data, such as semantic feature norms and abstractness, age of acquisition, and valence ratings. Finally, Semanticore will feature a blog showcasing a number of experts in semantic models and psycholinguistics, providing commentary and interesting data analyses on topics of interest. In this talk, we will demonstrate the many features of Semanticore.org and its usefulness to the psycholinguistic, cognitive Science, and machine learning community.

10:30 **Interactive Web-Based Graphics in Computerized Assessment**

Gary McClelland (gary.mcclelland@colorado.edu), University of Colorado at Boulder
Ulf-Dietrich Reips (u.reips@ikerbasque.org), University of Deusto; IKERBASQUE, Basque Foundation for Science

Computerized psychological assessments involving more than the most basic presentation of materials and using more than minimal response formats have required the use of technologies such as Java and Flash. Recent security problems and incompatibilities with these technologies on computers and their unavailability on many mobile devices have limited the use of more interesting modern computerized assessment techniques. We illustrate how the new drawing and data detection capabilities in HTML5 enable rich, engaging materials and response formats across computerized devices that have the promise of improving computerized psychological assessment. We will present examples for challenges with outdated

technologies and a variety of working examples in HTML5 and describe the construction of additional materials and response formats.

10:45 Using mobile phone applications to investigate mental arithmetic in everyday life

Oliver Lindemann (oliver.lindemann@uni-potsdam.de), University of Potsdam

Martin Fischer (martinf@uni-potsdam.de), University of Potsdam

Psychological research typically investigates human mental arithmetic under controlled experimental conditions, in which participants are required to focus for a longer period on the processing of a large amount of arithmetic problems. However, mental arithmetic under natural circumstances is characterized by an incidental processing of single isolated problems. It is therefore unclear to what extent findings obtained in the lab, such as the problem size effect, can be generalized. The present study aims to demonstrate how mobile phone applications can be used to get an insight into human numerical cognition in everyday life situations. We describe an Android app that replaces the default lock screen with a horizontal slider by a program depicting simple arithmetic problems together with multiple answer alternatives. Users have to select the correct solution to unlock and use their device. This app generated a large corpus of arithmetic problem solving speed and accuracy outside the lab. Here we report the analysis of performance of about 14000 subjects in about two million simple addition problems and discuss the results in the context of recent theories on mental arithmetic.

Session C: Nengo Tutorial (Sheraton D & E)

Nengo: A software suite for creating large-scale biologically realistic cognitive models

Terry Stewart and Chris Eliasmith, *University of Waterloo*

See Page 13 for details

11:00am-12:00pm Parallel Session 3

Session A3: Emerging Tools and Methodologies (Sheraton G)

Chair: Joe Houpt, *Wright State University*

11:00 **The Latest Systems Factorial Technology in R Developments**

Joseph Houpt (Joseph.houpt@wright.edu), Department of Psychology,
Wright State University

Leslie Blaha (Leslie.Blaha@wpafb.af.mil), 711th Human Performance Wing,
United States Air Force Research Laboratory

Devin Burns (devburns@indiana.edu), Psychological and Brain Sciences,
Indiana University

Systems factorial technology (SFT) is a powerful and mathematically rigorous framework for studying how cognitive systems use multiple sources of information together. Articles about SFT tend to focus on the mathematics and development of the theory, making them inaccessible many researchers. The SFT package for R was recently introduced to facilitate the use of SFT by a wider range of researchers. The original package contained tools implementing only the basic theoretical tools. In the last few years there have been a number of advances to SFT, which we will review, and we introduce their implementation in the SFT package. In particular, we will demonstrate R functions for functional principal components analysis of the capacity coefficient (Burns, Houpt & Townsend, 2013), calculating and plotting assessment functions (Townsend & Altieri, 2012), and calculating and plotting distributional bounds in a unified capacity space (Townsend & Eidels, 2011). Additionally, we expanded the package to include a function for the new capacity coefficient for single-target self-terminating (ST-ST) processing (Blaha, 2010), as well as functions supporting the plotting of cumulative distribution function bounds on the predictions of standard parallel processing models for minimum time, maximum time, and ST-ST decision rules.

11:15 **Assessing the Response Latency Variability of Very Large Touchscreens for Human Behavioral Research**

Leslie Blaha (Leslie.Blaha@wpafb.af.mil), 711th Human Performance Wing,
Air Force Research Laboratory

Andrew Derhammer (andrew.derhammer@gmail.com), University of
Akron; 711th Human Performance Wing, Air Force Research

David Dommert (David.Dommert@wpafb.af.mil), 711th Human
Performance Wing, Air Force Research Laboratory

Touch screens of all sizes, including phones, tablets, and table surfaces, offer

new options for both the display of stimuli and collection of response data for human behavioral research. However, because the display and response mechanisms are integrated, there are multiple sources of input timing whose variability should be assessed and accounted for in order to determine the device's contributions to the variability of human response time data, especially when millisecond precision is desired. We assessed input-to-software latency and end-to-end latency to understand the ability of the machine to first detect a touch input within standard software and then to respond with a visual cue (tactile-to-visual feedback). Multi-touch displays offer the further experimental advantage of handling multiple simultaneous user inputs, so the input-to-software and end-to-end measures were repeated for multiple simultaneous inputs. Because touch screens can include more complex interactions than a simple button press, we additionally assessed the variability of visual rendering under conditions where the user is dragging a finger across the screen. We applied our methodology to an 82 inch rear-projection Frustrated Total Internal Reflection (FTIR) Perceptive Pixel Multi-Touch Display, two 55 inch FTIR Perceptive Pixel Multi-Touch Displays, a 46 inch optical sensor Touch Table (model TT45), and a 32 inch reflected near-infrared Microsoft Surface table, all commercially available for laboratory use. We discuss the possible generalization of our methodology to the assessment of any touch screen device regardless of size to determine latency variability and suitability for response time experiments.

11:30 We're Watching You: Using Random Walks and Probability Trajectories to Profile System Users

Erica Snow (erica.l.snow@asu.edu), Arizona State University
Laura Varner (laura.varner@asu.edu), Arizona State University
Aaron Likens (aaron.likens@asu.edu), Arizona State University
G. Tanner Jackson (tannerjackson@asu.edu), Arizona State University
Danielle McNamara (dsmcnamara@asu.edu), Arizona State University

Research on individual differences indicates that students vary in how they interact with and perform while using intelligent tutoring systems (ITSs). However, relatively little work has been conducted to investigate how individual differences influence students' interactions with game-based features across time. The purpose of this study was to investigate students' patterns of interactions within a game-based intelligent tutoring system (ITS), and how those interactions varied as a function of students' reading ability. The analysis presented in this paper comprises a subset (n=40) of a larger study that included 124 high school students. Participants in the current study completed 11 sessions within iSTART-ME, a game-based ITS, that provides training in reading comprehension strategies. The current study employed a random walk algorithm and a transitional probability analysis to visualize student interaction patterns across time. These

analyses revealed that low and high reading ability students demonstrated distinctly different patterns of choices within the system. Specifically, low ability students' interaction patterns were anchored by one feature whereas high ability students demonstrated interactions across multiple features. These variations in interaction patterns shed light upon the various ways in which students engage in game-based environments. Specifically, they provide further evidence that individual differences play an important role in influencing students' interactions within learning environments. Examining users' paths may improve the adaptability of learning environments by affording researchers an opportunity to monitor and trace patterns of interactions as they evolve across time.

11:45 **Visual DMDX: a web-based authoring tool for DMDX**

Pablo Garaizar (garaizar@deusto.es), University of Deusto, Deusto Institute of Technology (DeustoTech)

Julen Telleria (1jutelle@opendeusto.es), University of Deusto, Deusto Institute of Technology (DeustoTech)

Ulf-Dietrich Reips (reips@deusto.es), Ikerbasque, Basque Foundation for Science

DMDX is a software package for the experimental control and timing of stimulus display for Microsoft Windows systems (Forster & Forster, 2003). DMDX is reliable, flexible, millisecond accurate and can be downloaded for free, thus it has become very popular among experimental researchers. However, setting up a DMDX-based experiment is burdensome due to its command based interface. Further, DMDX relies on RTF files where special keywords are used in a DMASTR-compatible syntax (Forster & Forster, 1975). Other experimental software such as E-Prime, Psychopy or WEXTOR became successful due to integrated graphical authoring tools. We thus decided to create Visual DMDX, a HTML5-based web interface to set up simple experiments and export them to DMDX item files format in RTF. Visual DMDX offers most of the features available from the rich DMASTR syntax, and it is a useful tool to encourage novice researchers in the use of DMDX. Moreover, we decided to release it under an open source license to foster collaboration in its continuous improvement.

Session B3: Semantic Modeling (Sheraton F)

Chair: Cyrus Shaoul, *Tübingen University*

11:00 **Evaluating a Method to Compare Human and Machine-Generated Semantic Spaces**

Xudong Huang (xhuang3@memphis.edu), University of Memphis

Yuchen Liao (yliao2@memphis.edu), University of Memphis

Jun Xie (jxie2@memphis.edu), University of Memphis

Jin Wang (wjn@memphis.edu), University of Memphis
Benjamin Nye (bdnye@memphis.edu), University of Memphis
Xiangen Hu (xhu@memphis.edu), University of Memphis

Many text corpus tools have been developed, but there is no gold-standard to select the best semantic spaces generated by these tools. We present a method called Induced Semantic Structure (ISS) that pragmatically answers the problem by measuring the differences between semantic spaces. While this does not guarantee an absolute best space, it can compare alternative spaces with various parameters (e.g. dimensions) and from various sources (e.g., algorithms, expert knowledge) based on nearest neighbor terms. The ISS method computes the overlap and order of the common nearest neighbors of a word across semantic spaces and outputs the ranking of the differences among spaces. By comparing spaces against human knowledge of a given domain (e.g. top neighbors of concepts in expert knowledge structure), a semantic space with the minimum difference can be identified as the best space available. This paper evaluates the effectiveness of the ISS method. The algorithm was validated by using ISS method to compute the differences between five Touchstone Applied Science Associates (TASA) spaces (3rd grade, 6th grade, 9th grade, 12th grade, and 1st college). TASA spaces are additive (i.e., 1st college corpus includes the corpora from all earlier grades). Therefore, spaces generated from corpora with the least overlap should have the greatest difference for an effective comparison measure. The ISS method accurately captured this pattern. This ISS method can be applied to compare algorithms for generating semantic spaces, compare computer-generated semantic spaces against human-generated ones, analyze social network trends, and build context-sensitive learning systems.

11:15 **A computational model of semantic convergence in bilingual lexicon**

Shin-Yi Fang (szf4@psu.edu), Pennsylvania State University
Barbara Malt (bcm0@lehigh.edu), Lehigh University
Eef Ameel (eef.ameel@psy.kuleuven.be), University of Leuven
Ping Li (pul8@psu.edu), Pennsylvania State University

Previous research by Ameel, Storms, Malt, and Sloman (2005) found that object naming patterns by bilingual speakers converge toward a pattern that is different from the naming patterns of either language, suggesting that bilingual lexical representations are not simply the sum of two separate monolingual representations. In this computational modeling study we investigate how cross-language lexical interactions come about in a connectionist model. We constructed a multi-layer self-organizing neural network to simulate the learning processes for lexical representation. The Self-Organizing Maps (SOMs) were used for the representation and organization of lexical semantic and object information. The associative

links between and within SOMs are updated by Hebbian learning. SOM relies on unsupervised learning algorithm (Kohonen, 1995) to identify input similarities in the high-dimensional space and derives a topologically ordered input structure on 2-D space. We compared our modeling results with Ameel et al.'s empirical data from bilingual contexts. The model has successfully simulated that bilinguals develop a pattern of convergence that matches with empirical data reported in Ameel et al.. Furthermore, our modeling data suggest that bilingual convergences are partly based on the strength of name agreement in both L1 and L2 monolinguals, and that there are bidirectional influences between L1 and L2 naming patterns. The weaker the name agreement in L1, the more likely an object category is to be influenced by L2. The cooperation and competition of object vs. naming categories between the bilingual's two languages play an important role for semantic convergences in bilingual lexical representation.

11:30 An R package for large-scale naive discriminative learning.

Cyrus Shaoul (cyrus.shaoul@uni-tuebingen.de), University of Tuebingen
Petar Millin (petar.milin@uni-tuebingen.de), University of Tuebingen
Peter Hendrix (petrus.hendrix@uni-tuebingen.de), University of Tuebingen
Michael Raschmar (michael.raschmar@uni-tuebingen.de), University of
Tuebingen
R. Harald Baayen (harald.baayen@uni-tuebingen.de), University of
Tuebingen

Abstract: We introduce naive discriminative learning (NDL) model, based on the Rescorla-Wagner equations, and simulated using the equilibria approximation proposed by Danks (2003). NDL allows the consequences of accumulated experience in lexical processing to be explored with realistic data samples. This means learning to use thousands of cues to discriminate tens of thousands of lexemes, with training on corpora ranging from tens of millions of words to several billions of words. However more training data is not necessarily better. For example, from an information-theoretic perspective, entropy increases with vocabulary/corpus size, and the model explains how processing speeds inevitably slow down as experience grows, such that the model fits empirical aging data with no "decline". We also present results from recent work predicting eye-movements and EEG waveforms using weights derived from our NDL models. Our implementation of NDL is a package that is freely available for the open-source R statistical computing environment.

11:45 Evaluating topic trend detection in online text streams

Jun Xie, University of Memphis (jxie2@memphis.edu),
Yuchen, Liao, University of Memphis
Xudong, Huang University of Memphis

Yun Wu, University of Memphis
Jin Wang, University of Memphis
Benjamin Nye, University of Memphis
Xiangen Hu, University of Memphis

Online text streams, such as blogs and Twitter, play a key role in popular opinion formation and trend detection. Compared to traditional texts (e.g. scientific papers, books), online text streams are dynamic information arriving continuously over time. Therefore, opinion detection for online text streams not only considers the relevant content for a given topic but also how the topic evolves (i.e., the topic trend). A moving window, which is a time frame sliding along the time interval of interest, enables real-time online computation of topic trends. However, existing methods often track trend in offline manner. Furthermore, the moving window size may influence semantic algorithms' performance on trend tracking. This paper proposes a simple online approach to track topic evolution within domains and measures the impact of moving window sizes on performance. The corpus to be analyzed uses Weibo posts, which are similar to Tweets and extremely popular in China. Moving windows with different size are being used to segment corpus for the chosen topic evolution tracking. Then, for each type of moving window, Latent Semantic Analysis builds temporal semantic spaces to generate topic revolutions. Results find that topic evolutions in different domains show diverse patterns, and moving windows with different size have different capabilities to capture real-time topic evolutions within domains.

Session C: Nengo Tutorial (Sheraton D & E)
Continued from previous session
See Page 12 for details

1:00-2:30pm Poster Session, Sheraton Hall

Poster Presenters: Please follow the number designation for your poster setup. Poster must be set up between 12:00-1:00pm and must be removed by 3:00pm. Presenters of odd-numbered posters are to be at their posters for the first 45 minutes, and presenters of even-numbered posters at theirs for the second 45 minutes of the poster session.

1. Speech Act in Epistemic Game: In and out of the Context

Haiying Li (hli5@memphis.edu), University of Memphis
Danielle Clewley (dnclwley@memphis.edu), University of Memphis
Brea Walker (bswlker2@memphis.edu), University of Memphis
Zhiqiang Cai (zcaai@memphis.edu), University of Memphis
Arthur Graesser (graesser@memphis.edu), University of Memphis

Abstract: Speech act category acts as one of the indispensable elements to detect the speaker/writer's intention. The state transition of speech act for online chats in epistemic game provides the conversational patterns and lays foundation for the development of the conversational computer agent. In the epistemic game "Land Science", a group of players develop land use plans as interns at an urban and regional planning firm. A mentor guides the players to accomplish the related tasks by chatting with the players. Meanwhile, players collaboratively solve the problems by chatting during the game. The state transition network of speech act during online chats in the game is essential to detect the conversational patterns among players and human mentor. This study first developed a more practical taxonomy of the speech act and then manually annotated the data set either in the context or independent of the context. We performed a disagreement and difficulty analysis to identify the bottlenecks of our annotating procedure. The ultimate objective is to design an automatic speech act classification mechanism for online chats, to construct the state transition network of speech act in epistemic game, and finally develop the conversational compute agent for the epistemic game. This study is important for this purpose. The results indicated that overall speech act annotation is more reliable out of the context (Average Kappa = .66) than that in the context (Average Kappa = .55) with the highest reliability for greeting, question, and metastatement.

2. Do Students in Developmental Reading Literacy Programs Process Text Differently than Traditionally Admitted Students?

Greta Chan (gchan@niu.edu), Northern Illinois University
Melissa Ray (mray4@niu.edu), Northern Illinois University
Sonya Armstrong (sarmstrong@niu.edu), Northern Illinois University
Joe Magliano (jmagliano@niu.edu), Northern Illinois University

Many students enrolled in universities and community colleges are under-prepared to meet expectations for reading literacy in their courses. As such, many students are referred to remedial programs, called developmental education (DE) programs, to bolster their reading literacy skills. Referral criteria are usually based on standardized tests. However, these tests do not provide information about what students do in the process of reading. In this exploratory study, we used a computer tool, the Reading Strategies Assessment Tool (RSAT), to assess the mental processes that are involved in the course of reading. Our goal is to find out in what ways students in DE programs process the text differently from students enrolled through traditional admittance criteria. RSAT analyses revealed DE students were less likely than non-DE students to relate the sentence that they were reading to the background knowledge that supported the comprehension of that sentence. On the other hand, these data suggest that DE student engage in some inference processes important for comprehension similarly to their non-DE peers. Our data do not only identify mental processes that distinguish struggling readers from a comparison group, but also provide a proof of concept for the utility of RSAT as an assessment tool for some remedial programs.

3. **Evaluation of Google Translation at Multi-textual Level**

Haiying Li (hli5@memphis.edu), University of Memphis
Arthur Graesser (graesser@memphis.edu), University of Memphis
Zhiqiang Cai (zcaai@memphis.edu), University of Memphis

Abstract: This study aims to evaluate the quality of Google English translation at multi-textual level, providing the foundations for descriptions and understanding the quality of Google translation. This study extends the restriction of the current evaluation of the Google translation at the phrase or sentence level to the multi-textual level, including narrativity and cohesion. The automated text analysis tools used in this study encompassed Chinese and English Linguistic Inquiry and Word Count (LIWC), Latent Semantic Analysis (LSA), and Content Word Overlap (CWO). Component model was used to compute component scores in terms of analysis results from LIWC, LSA and CWO. Meanwhile, cohesion score was computed from LSA and CWO. The original Chinese sample was a collection of 289 speeches and writings composed by the Chinese political leader Mao Zedong. The quality of Google translation was evaluated by two types of comparisons. The first comparison was the comparison of the accuracy of genre, cohesion and other discourse and psychological features between Google translation and human translation. The second comparison was the accuracy of both English translations with the original Chinese version at these three levels. The results showed that Google English translation was significantly correlated with human translation. Moreover, the results also showed Google translation was more similar than human translation to original Chinese.

4. **Measuring the trait component of speech disfluency using computerized language measures**

Sean Murphy (smurphy1@gmail.com), New York Psychoanalytic Society and Institute

Bernard Maskit (daap@optonline.net), Mathematics Dept. Stony Brook University

Wilma Bucci (wbucci@optonline.net), Derner Institute, Adelphi University

Disfluency is a characteristic feature of spontaneous spoken language that is associated with speech production functions and problems. Indicators of this dimension include repetitions, false starts, repairs and lexical types, such as 'uh', 'um', and 'like.' Variation in these indicators have been investigated either separately or in combination in relation to mechanical speech planning difficulty, management of listeners' expectations, deceptive speech, and in other contexts. This study used measures of disfluency computed by the Discourse Attributes Analysis Program (DAAP) (thereferentialprocess.org) to investigate the trait variance of disfluent speech in monologue and dialogue samples. The monologue sample included narratives of 63 college students who told their earliest memory, a life event and a story in response to a picture on two occasions six weeks apart. The dialogue sample was taken from the Switchboard Corpus of telephone conversations where each of the 99 speakers in the sample were matched with a different conversation partner on each of two occasions six weeks apart. In the monologue sample, overall disfluency scores were significantly correlated over time ($r=.47$ to $r=.65$) for each of the three narrative tasks. In the dialogue sample, speakers' disfluency correlated $r=.64$ with themselves over the six week period, while the correlations with their conversation partners were either much lower $r=.26$ or not significant $r=.06$. These results indicate considerable trait variance in disfluency and support the use of within subjects designs in studies of factors affecting variation in disfluency.

5. **Examining noun-bias in Spanish using the CHILDES database**

Meaghan Mauer (mauerm@emmanuel.edu), Emmanuel College

Xiaowei Zhao (zhaox@emmanuel.edu), Emmanuel College

Study of various languages is important to broaden the understanding of child language development and acquisition. The Noun-bias hypothesis predicts that nouns are acquired earlier in children's language development than other word categories, and children's early vocabulary includes more nouns than verbs or other word categories. Previous research has shown strong support for noun-bias in the language development in English (Gentner, 1982). In the present study, we address the question of whether noun-bias is cross-linguistic in nature. We examine the lexical compositions of several lexical

categories in Spanish speaking children's vocabularies across eight age groups ranging from 13 to 60 months. Our analyses are based on data obtained from the Child Language Data Exchange System (CHILDES, MacWhinney, 2000). We examined data from 14 corpora containing transcripts of native Spanish-speaking children's speech under various linguistic contexts (in homes, in schools, or in labs). We grouped transcripts based on children's age and analyzed word frequencies for the vocabularies of each age group using the CLAN program. This information is used to calculate noun-verb ratios of children's speech at each age group to examine the existence of a noun-bias in Spanish language acquisition. The results are further compared with other languages like English and Chinese.

6. **Exploring the role of stimulus-response compatibility on the spatial Stroop effect**

Min-Ju Liao (mjiao@ccu.edu.tw), National Chung Cheng University
Shang-Hung Wang (ry3298@gmail.com), National Chung Cheng University

The present study investigated the compatibility of stimulus codes and response modalities and how they mediated the spatial Stroop effect on an interface display. Stimulus code (word vs. arrow vs. moving dots), response modality (voice vs. keypress vs. mouse movement), 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 make a left or right response. Spatial Stroop interference occurred when the directional information of the stimulus was incongruent with its physical location. Word-voice, arrow-keypress, and arrow-movement appeared to be more compatible pairings than the other combinations. For the word stimulus, incongruent spatial location delayed the reaction time for all response modalities and resulted in longer movement path length for the mouse movement. For the moving dots, incongruent spatial location speeded up the vocal and keypress responses. No effect was found for the arrow. The present study showed that the direction and magnitude of the spatial Stroop effect were mediated by the stimulus code, response modality and their relative compatibility. Arrow was recommended for conveying directional information on an interface when spatial interference is concerned.

7. **Crossword Expertise as Recognition Decision Making: An Artificial Intelligence Approach**

Kejkaew Thanasuan (kthanasu@mtu.edu), Michigan Technological University
Shane Mueller (shanem@mtu.edu), Michigan Technological University

Solving crossword puzzles incorporates two important aspects of lexical memory search: semantic clues that indicate the meaning of the answer, and orthographic patterns that constrain the possibilities. Expert solvers are

extremely fast at accessing the vast set of information available, making them a useful analog to many expert decision domains. Furthermore, large clue corpora exist and thus enable models of their expertise to be populated with vast amounts of real knowledge, unlike most domains of expert study. In this study, we will compare expert and novice human solvers to a computational simulation model of recognition-based decision processes to evaluate the relative importance of orthographic and semantic memory search in supporting crossword expertise. Results reveal that although experts have superior orthographic-based memory search, expert solutions are much more highly determined by breadth of semantic knowledge and the semantic fluency of generating these responses. These effects are explored via an AI model using past crossword clues as a basis for its knowledge base.

8. **The Development of An Automated Analyzer for Chinese Text**

Jih-Ho Cha (toddcha@gmail.com), National Taiwan Normal University
Ju-Ling Chen (chenjuling@ntnu.edu.tw), National Taiwan Normal University
Tao-Hsing Chang (thchang@cs.nctu.edu.tw), National Kaohsiung University of Applied Sciences
Yao-Ting Sung (sungtc@ntnu.edu.), National Taiwan Normal University

Chinese readability index explorer (CRIE) is the first automated Chinese text analyzer whose core modules and technologies include segmentation, parser, writing corpora, Latent Semantic Analysis, Support Vector Machine (SVM) and other components widely used in computational linguistics. CRIE provides linguistic information, readability prediction, and writing diagnosis for Chinese native speakers as well as reading materials for Chinese as a Foreign Language (CFL) learners. Based on the idea that comprehension involves multi-level cognitive processes, CRIE implements 4 levels (word, syntactic complexity, semantic relation, and cohesion) of 75 features to cover a wide range of linguistic and discourse representations of a text. CRIE incorporates SVM to make readability level predictions because of its better capability of processing linearly non-separable and not normally distributed data. The prediction accuracy of texts for native Chinese speakers using grade levels as criterion reaches 71.75 %. The prediction accuracy of texts for CFL using the 3 levels of CEFR as criterion reaches 86%. CRIE also provides diagnostic reports which offer error analysis, and revised suggestions.

9. **Collecting confidence ratings in cognitive psychology experiments: Effects of scale and entry method**

K. Andrew DeSoto (desoto@wustl.edu), Washington University in St. Louis

Many researchers investigating cognitive or metacognitive research questions collect retrospective confidence ratings in their studies. These ratings indicate an individual's certainty that he or she responded to a memory decision

accurately, and often are correlated with performance or examined in their own right by researchers. Although confidence ratings are commonplace in empirical studies, the methods by which they are collected vary from researcher to researcher, and little to no theoretical backing underlies commonly employed practices. To explore further the intricacies in collecting confidence ratings, I present confidence data from several studies (published and unpublished) that provide evidence that the method in which confidence ratings are collected can affect the conclusions that are drawn. Specifically, ratings collected using a numeric keypad are subject to biases of rounding (i.e., an apparent preference for round numbers), and ratings collected using a graphical slider suggest a habitual preference toward the default value. The implication is that the way in which confidence ratings are collected must be considered carefully before researchers embark on a project involving confidence.

10. **Using SuperLab to Improve Understanding of Research Design**

Chris Koch (ckoch@georgefox.edu), George Fox University

Psychology majors often do not have technical experience designing computer-based experiments. While there may be practical reasons for encouraging students to develop experiments using a programming language or experimental software package, an important pedagogical question is whether or not doing so increases student understanding of experimental design. Students were required to develop an experiment using SuperLab in an advanced research methods class. Research topics were chosen individually by students. Experiments were evaluated with a rubric common to computer science programs. Students indicated that they had a better understanding of how to design and deliver an experiment as a result of completing the SuperLab assignment. Furthermore, they indicated that they understood how to organize and efficiently create an experiment using SuperLab. Learning how to properly develop an experiment, however, is limited in its transfer. For instance, developing an experiment did not necessarily help students understand the data produced from the study. Therefore, teaching how to develop computerized experiments can provide a context for a larger review of research-related skills including data file management and statistical analysis. Interestingly, when asked an open-ended question at the end of the course, 10 of the 17 students indicated that they would like to spend more time designing additional experiments, more time using SuperLab, and more time working with data files. These responses are consistent with the idea that developing computer-based experiments increasing interest in designing experiments and a context for working on additional research-related skills.

11. **Typicality and Age of Acquisition across Sensory Domains**

Brent Kievit-Kylar (bkievitk@indiana.edu), Indiana University
Michael Jones, Indiana University
Peter Todd, Indiana University
Yong-Yeol Ahn, Indiana University

We explore the relationship between typicality as measured across and within sensory domains and the relationship of Age of Acquisition with these typicality measurements. We focus on two categories of words --- fruits and vegetables --- that are present both in the Rosch typicality norms and in our nutrition dataset. We discovered the average sensory representation of all items within each category (as a vector over a multi-dimensional space) and defined a similarity metric between each item and this average representation (which we will define as a Similarity to Average measurement or StoA). Comparing each item's StoA with its typicality, we determined striking differences between typicality, domain of similarity, and category. In particular, we found that odor was a strong predictor of fruit typicality, visual appearance a strong predictor of vegetable typicality and health a strong predictor of the typicality measures in both categories. We also compared the StoA of these words to their Age of Acquisition (AoA). We determined that in each sensory domain, a quadratic curve better fit this relationship than a linear one. Also, in each sensory domain, the curve took a similar shape with a higher average AoA for both low StoA and high StoA words, but lower AoA for medium StoA words. This indicates that words that are too unusual (relative to the average) or too common and similar (close to average) are actually learned later in life as compared to those that reside in the middle of the space. A similar pattern was identified for cross-modal prediction where words that were not too easy or too hard to predict between domains (a measure of similarity across domains) were learned earlier.

12. **Website with Helpful Resources for Learning How to Conduct Research via the Internet**

Michael Birnbaum (mbirnbaum@fullerton.edu), California State University, Fullerton

This poster advertises a free Website providing instructional materials used to teach researchers about how to conduct behavioral research via the Internet. The Website is freely available and has links to materials used by several instructors in a series of advanced training institutes, supported by the National Science Foundation and held on the campus of California State University, Fullerton over the last dozen years. Many working examples are freely available via the Website, illustrating materials on HTML forms, CGI scripts, stimulus construction and delivery, JavaScript programming for dynamic, interactive experiments, and other techniques. Special programs

developed to assist experimenters to construct Web-based studies and experiments are also linked or described there. This Website will continue to exist for at least the next two years. Those interested in learning about these techniques can find helpful materials at this site, which contains links to other resources available via the Web.

13. **Accounting for the nature of human response selection in evaluating semantic models**

Timothy Rubin (timrubin@indiana.edu), Indiana University
Michael Jones (jonesmn@indiana.edu), Indiana University

A standard approach within computational linguistics involves fitting semantic models to human responses on tasks such as multiple-choice synonym tests from the TOEFL or ESL examination. Fitting semantic models to this type of data plays an important role in semantic model development, tuning, and comparison. Despite the importance of these tasks to semantic modeling, oftentimes they do not account fully for the nature of human response data involved in the task. In particular, multiple-choice synonym tasks are evaluated in terms of the percentage of correct responses achieved by the model, where the correct response is determined by the company that developed the test. However, human responses on these questions comprise a distribution over response choices that reflects the relative similarity between these choices and the target word. In some cases, these response distributions indicate that there is high confusability between the response alternatives, or even that one of the "incorrect" responses is more semantically similar to the target word for the plurality of responders. Given that the goal semantic modeling is to capture human cognition and behavior, it follows that semantic models should capture the variation of human responses across different questions. Here we look at the use of alternative evaluation metrics--such as rank-correlation measures and log-likelihood of fit using a Luce choice rule--that account for the similarities between the target word and all response-choices. We show that this improved approach for model evaluation can significantly affect an observed model's performance.

14. **Study of the communication strategies used by Australian CFL learners in Second Life**

Yu-Ju Lan (yujulan@gmail.com), National Taiwan Normal University

The current study aimed at confirming the potential of Second Life (SL), a 3-D multi-user virtual environment released by Linden Lab in 2003, for promoting the oral output of learners of Chinese as a second language (CSL). The study applied an experimental research design. Twenty CSL learners participated in this study: 10 of them were randomly assigned to the control group, and the others to the experimental group. During the experiment, all the participants

received identical learning materials except the learning environments. The CSL learners of the control group learned in a conventional environment, whereas the experimental group learned in SL. All the teaching/learning processes were recorded and analyzed to investigate the differences in oral output by the two groups of CSL learners. The analytical results showed that the CSL learners of the experimental group significantly produced more open-ended oral output than those of the control group did. In sum, the findings of the current study confirmed the potential of SL for promoting in-class oral output by CSL learners.

15. **Recycled Sas PrEdiCTions (RESPECT)**

Ron Hays (drhays@ucla.edu), UCLA
Karen Spritzer (spritzer@ucla.edu), UCLA

Recycled predictions are used to understand the marginal effect of independent variables on a dependent variable (Basu, 2005; Graubard & Korn, 1999). They are obtained from regression models by averaging predicted scores on the dependent variable after fixing the value of one independent variable (e.g., setting gender to be “female”) and using observed values on the remaining independent variables in the sample. Li and Geeta (2010) presented a SAS macro to obtain recycled predictions for logistic regression. SAS users can obtain recycled predictions for ordinary least squares regression models using SAS-callable SUDAAN code (Bieler et al., 2010), but this requires purchase of a SUDAAN software license. This paper summarizes the use recycled predictions for ordinary least squares regression models and describes a macro that uses SAS PROC GLM to obtain them.

16. **Emoticons Behaving Ambiguously**

Sarah Maples (smapl001@ucr.edu), University of California, Riverside
Curt Burgess (curt@ucr.edu), University of California, Riverside

A semantic priming experiment was conducted with word primes and emoticon targets. The strength of word-emoticon relationships was determined by a norming study. There were three word – emoticon relationships, e.g., primary, happy :) or secondary, blushing :) or unrelated, bitter :). Primes were presented after a fixation point for 500 ms, followed by the emoticon target. Subjects made semantic relatedness judgments. Response times to emoticon targets preceded by a primary meaning were faster than unrelated targets. However, response times to secondary meaning targets were longer than the unrelated targets. These results are discussed from a multiple meaning activation viewpoint (i.e., Simpson & Burgess, 1985) and in the context of high-dimensional memory models such as HAL (Burgess & Lund, 2000).

17. **The Use and Connotation of the N-word in Twain's Novels and in Contemporary Society**

Curt Burgess (curt@ucr.edu), University of California, Riverside
Peter Cao (pcao002@ucr.edu), University of California, Riverside

The use of the N-word is viewed by most as an egregious utterance with strong racist meaning. Current controversy exists that has resulted from school boards and parents wanting to remove texts such as Tom Sawyer and The Adventures of Huckleberry Finn from school libraries or replace the texts 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. In order to test the semantic hypothesis that the word was used in a neutral sense in the early literature, a semantic model was built using the texts from Tom Sawyer and The Adventures of Huckleberry Finn. The results of this were compared to a model built from USENET text from the 1990's. The comparison was made by an analysis of the semantic neighborhoods of various forms of the N-word and the word slave. The word slave was used infrequently in the Twain novels so an analysis of its usage is compared to the 1990's neighborhood results. The results are consistent with the reported etymological history and the range of the contemporary usage was extensive. A further analysis of semantic distances between the word slave and a set of negative and positive emotionally valenced words suggests certain problems inherent in the substitution of the word slave for the N-word.

18. **The Validity of Osgood's Emotional Word Features**

Dawn Carroll (dcarr010@ucr.edu), University of California, Riverside
Priscila Mendoza (pmend010@ucr.edu), University of California, Riverside
Jacklyn Kozich (jackiekozich@netzero.net), University of California, Riverside
Justin Estep (jeste002@ucr.edu), University of California, Riverside
Curt Burgess (curt@ucr.edu), University of California, Riverside

In 1962, Charles Osgood argued that emotional aspects of meaning are linguistically universal. He characterized these emotional dimensions of words as Evaluation, Potency and Activity. Heise (1965) acquired extensive normative information on over 1,500 words on these three emotional features. For example, Beauty was rated as strongly positive (Evaluation) and Potent, but less so on Activity. The word Victory was similarly positive, but more Potent and Active than Beauty. The word Hunger was rated as strongly negative, very Potent, but was rated low on Activity. These dimensions, first discussed by Osgood, remain of considerable interest. Heise (2001) has developed an online method of acquiring the ratings and is employing this with different languages and cultures. The results presented here evaluate Osgood's theory and Heise's normative data by using a computational

semantic model of word meaning. A potential limitation of the normative data is that the three dimensions are all inter-correlated (0.40 to 0.47). This presentation treats the emotional dimensions as categories and the word vectors generated by the computational model are used in a multi-dimensional scaling solution and word frequency is used as an additional variable in these analyses.

19. Boring Begets Ignoring: The Emotions of Learning with an Intelligent Tutoring System

Audrey Weil (weilam@miamioh.edu), Miami University
Christopher Fisher (fishercr2@miamioh.edu), Miami University
Christopher Wolfe (wolfecr@miamioh.edu), Miami University

We analyzed videos of five participants interacting with BRCA Gist, an Intelligent Tutoring System built on the AutoTutor Lite platform. BRCA Gist uses artificial intelligence techniques such as Latent Semantic Analysis to engage users in tutorial dialogues with a talking avatar. It presents didactic information using screen text, images, sounds, and video clips. BRCA Gist teaches users about genetic breast cancer risk, and the complicated issues involved with genetic testing. It is comprised of 25 segments discussing different issues related breast cancer risk and testing. The entire tutorial, including 7 tutorial dialogues, takes approximately 90 minutes. We coded each segment of video clips for the emotions of learning, specifically boredom, confusion, flow, frustration, and anxiety. Previous research suggests that these are the affective states associated with learning. This small study was embedded in a larger study on the efficacy of BRCA Gist that provided clear evidence that BRCA Gist participants performed significantly better than two comparison groups. Nonetheless, we found that the most prevalent emotion was boredom. Moreover, we found a positive correlation of .79 between the frequency of boredom and the number of segments participants skipped during the tutorial. The participants who showed the most boredom also performed quite poorly on knowledge tests at the end of the tutorial. We found little evidence for flow, frustration, and anxiety. Results are discussed in the context of Intelligent Tutoring Systems, the prospect of providing them with the capacity for automatic emotion detection, and approaches to making BRCA Gist more engaging.

20. A Wii bit of help: Cooperative gaming increases number of subsequent spontaneous helping between the players

Igor Dolgov (id@nmsu.edu), New Mexico State University
Matthew Nearents, New Mexico State University
William Graves, New Mexico State University
C. Brooks Volkman, New Mexico State University
Jeremy Schwark, New Mexico State University

Cooperative gaming is the preferred form of entertainment among children and teens. Although video game content is typically violent, often producing negative social outcomes, cooperative game play ameliorates its anti-social impact. We examined the influence of cooperative/competitive gameplay on subsequent spontaneous helping and the mitigating role of customized avatars. Participants played cooperatively or competitively with a confederate; the game was Wii doubles tennis in Experiment 1 and Wii canoeing in Experiment 2. Results revealed that, in both experiments, participants who cooperated picked up significantly more pens spilled by confederates after gameplay than those that competed. This effect held true regardless of avatar customization in Experiment 2, whereas only those that customized their avatars in Experiment 1 subsequently helped more. Our findings support the few recent publications which suggest that game structure, particularly whether a game involves cooperation, may have greater ramifications for future player behavior than game content. Despite the reality that most games are aggressive in nature and that exposure to this kind of media has serious negative societal outcomes, the worldwide emergence of multiplayer gaming may have unexpected prosocial benefits.

21. **The Use of Explanation and Argumentation in Intelligent Tutoring System Dialogues**

Elizabeth Cedillos (cedillem@miamioh.edu), Miami University
Christopher Wolfe (wolfecr@miamioh.edu), Miami University

We tested the effects of student-generated arguments and explanations on learning in the ill-structured domain of photography using the intelligent tutoring system AutoTutor Lite. Participants (118) were randomly assigned to one of four conditions: didactic tutorial with argumentation interaction, didactic tutorial with explanation interaction, didactic tutorial only, or an irrelevant control tutorial. The didactic tutorial gave a short lesson on photography. After completing the didactic tutorial participants were given a multiple-choice knowledge test, made photography judgments on 6 different dimensions of photography, and gave advice on four photography scenarios. Results show a significant difference in knowledge item scores $F(3, 116) = 18.27, p < .001$. Tukey's HSD post-hoc test indicates all didactic tutorial conditions performed better than the control, however, there were no significant differences in scores among the didactic-argument, didactic-explanation, and didactic only conditions. In the photography judgment task, there were significant differences in 2 of the 6 dimensions of photography $F(3, 115) = 2.91, p < .05$ and $F(3, 114) = 4.81, p < .01$. Post-hoc tests indicate that for the dimension "composition," the didactic only condition performed significantly better than the control. Additionally, for the dimension "shutter speed," the didactic only condition outperformed the didactic-explanation condition, which in turn outperformed the control condition. We present

qualitative data on photography advice in the context of argumentation dimensions such as claims, warrants, reasons, and backing. Results are discussed in terms of theories of self-explanation and argumentation in an ill-structured aesthetic domain.

22. **Probabilistic negation: computer mouse-tracking reveals non-linear distortions of truth perception when negated**

Stephanie Huette (shuette@memphis.edu), University of Memphis
Sarah Anderson, Nielsen
Michael Spivey, University of California, Merced

Abstract: Previous work has shown that a continuum of truth is reflected in real-time motor movement behavior (McKinstry, Dale & Spivey, 2008). In a mouse-tracking paradigm, participants responded yes or no to statements of varying truth-values such as "A thousand is more than a million" or "English is a language" as well as more ambiguous statements such as "Murder is sometimes justifiable". In the present study, we replicated these results along an 11-point continuum of truth-values, finding that the end-points of averaged mouse trajectories vary as a function of truth-value. In addition to this, negated versions of each stimulus were tested and revealed that truth-values for negated sentences follow more complex trajectories and do not preserve the original truth-value of the statement. The evidence found presents a problem for theories of negation that require a revision from the affirmative meaning. Alternative mechanisms for how truth is affected by negation are proposed.

23. **VCAEST: Training Facilitated by an ITS embedded in a Virtual World**

Keith Shubeck (kshubeck@memphis.edu), University of Memphis
Mae-Lynn Germany-Shubeck (mlgerman@memphis.edu), University of Memphis
Scotty Craig (Scotty.Craig@asu.edu), Arizona State University
Parvati Dev (parvati@clinispace.com), CliniSpace
Xiangen Hu (xhu@memphis.edu), University of Memphis
Robert Koch (rakoch@memphis.edu), University of Memphis
Wm LeRoy Heinrichs (wlh@innovationinlearning.com), Innovation in Learning, Inc.

Live-action training simulations supported by expert facilitators are an expensive approach to learning with logistical challenges. Facilitator-supported learning may not adequately address the individual needs of learners, particularly when training must be conducted in large groups. Intelligent tutoring systems (ITSs) have been shown to provide effective alternatives to tutoring and training experiences. Virtual worlds provide unique opportunities for distance training which can emphasize situational

awareness and team-oriented task completion. The Virtual Civilian Aeromedical Evacuation Sustainment Training (V-CAEST) interface takes advantage of both virtual worlds and ITSs in an effort to provide an accessible, inexpensive, and robust training environment for civilian medical personnel. V-CAEST aims to improve communication between civilian and military medical personnel during disaster situations. V-CAEST is modeled after the live-action training program, CAEST, and serves as a virtual implementation of these training simulations. V-CAEST integrates the web-based ITS, AutoTutor LITE (Learning in Interactive Training Environments), into a virtual world. Here, AutoTutor LITE guides, educates, and provides feedback to users in real time. AutoTutor LITE incorporates Natural Language Processing enabled by a Domain Specific Semantic Processing Portal (DSSPP), which allows the system to automatically evaluate student responses. This DSSPP takes in military and medical corpora to create a domain-specific semantic space. This presentation will cover the advantages and challenges of incorporating an ITS into a virtual world. This presentation will also propose a methodology to empirically validate the marriage of these two technologies by reporting usability data and on the progress of a randomized trial between V-CAEST and CAEST.



PSYCHOLOGY
SOFTWARE TOOLS



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



E-Prime
Extensions for
tobii



E-Prime Extensions for fMRI