SCiP 2011 Program

Seattle, Washington
Thursday, November 3 2011

Message from the SCiP President

Welcome to the 41st annual meeting for the Society for Computers in Psychology (SCiP). The purpose of SCiP is to increase and disseminate knowledge of the use of computers in psychological researcher and educational endeavors. We have an excellent program this year that certainly fulfills this purpose. You will see several papers that present new computer-based research and educational tools, as well as novel uses of existing approaches to address content issues. I want to thank each of the presenters and our keynote speaker, Susan Dumais, for taking part in the conference. We obviously couldn't do it without you.

I want to draw your attention to one change this year. I have decided to organize a Presidential Symposium rather than giving the traditional, Presidential Address. As some of you may know, over the years I have presented research focusing on developing natural language processing approaches for educational applications. SCiP has certainly presented an important venue for me to present this research and get feedback from my peers. However, it has also been an important venue for me to learn about the work of other researchers in this area. And so, I wanted to organize an event that would reflect researchers who have been influential in this line of research. I'm very excited about the symposium and I hope you find it informative.

I have been regularly attending SCiP for a little over 10 years and over this time, I have developed a strong sense of commitment to this community. Part of that sense of commitment came with a growing understanding how important SCiP has been to experimental psychology – which continues today. Sure, we can buy an experimental package to run our experiments, but there are always new challenges in using new technologies, statistical approaches and packages, computational modeling, etc. Each year I have learned something new that I can incorporate in to my program of research. I'm certain than many of the regular attendees can say the same.

If you are new to SCiP, welcome and I hope you enjoy the conference. If you have been here before, welcome back. Regardless of whether you are new or old to the society, you have a voice in SCiP. If I've learned anything about SCiP over the years, it is that we all make a difference and are important to its success.

Enjoy SCiP 2011 and your visit to Seattle!

Joe Magliano

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 microcomputers 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

SCiP 2011 is organized by the SCiP offcers and steering committee members listed below. Please contact them if you have any problems during the conference.

- President: Joseph P. Magliano
  Northern Illinois University
  jmagliano@niu.edu
- President Elect: Ping Li
  Pennsylvania State University
  pul8@psu.edu
- Past President: Michael H. Birnbaum
  California State University, Fullerton
  mbirnbaum@fullerton.edu
- Secretary-Treasurer: Michael Jones
  Indiana University
  jonesmn@indiana.edu

Steering Committee

- Patrick Conley
  University of Wisconsin
  pconley@uwsp.edu
- Otto MacLin
  University of Northern Iowa
  Otto.MacLin@uni.edu
- Giuseppe Riva
  Università Cattolica del Sacro Cuore
  http://www.giusepperiva.com/
- Danielle McNamara
  University of Memphis
  d.mcnamara@mail.psyc.memphis.edu
- Douglas Mewhort
  Queen's University
  mewhortd@queensu.ca
- Kay Livesay
  Linfield College
  klivesa@linfield.edu
- Randy Jamieson
  University of Manitoba
  jamiesor@cc.umanitoba.ca
- Rick Dale
  University of Memphis
  rdale@ucmerced.edu
- Yana Weinstein
  Washington University St. Louis
  y.winstein@wustl.edu

Conference Program

The conference program contains the conference schedule, the abstracts of all presentations and a list of authors. Please bring the program with you each conference day. Printouts of the conference schedule and maps of Seattle are available at the secretarial desk in the central hall free of charge.

The hotel location is shown in the map.

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.

Figure 1: Sheraton Seattle Hotel

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 3rd, from 8:00 till 9:30.

Key Note

Dr. Susan Dumais, from Microsoft

"Understanding in situ search behavior using large-scale log analysis and web-based experiments"
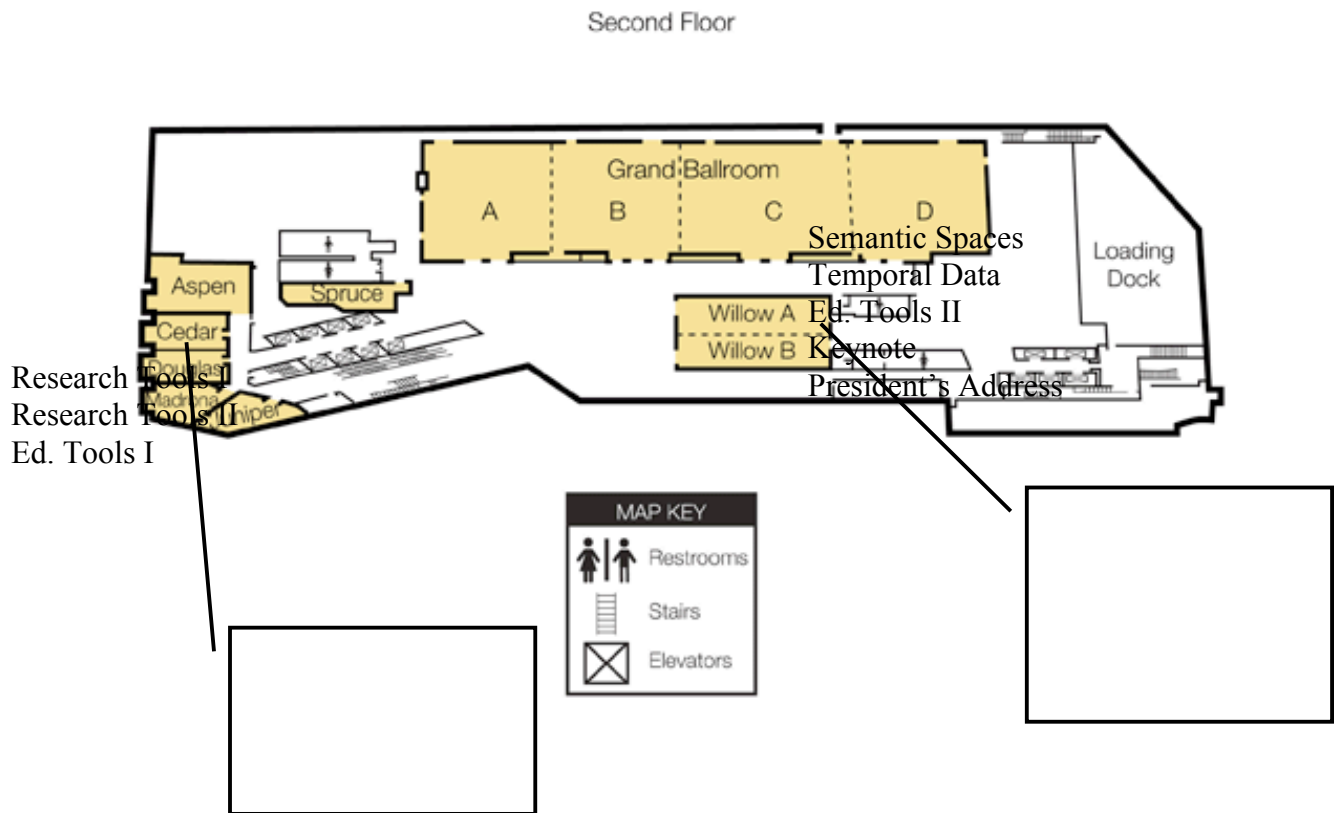
Location
Sheraton Seattle Hotel

The conference will be held at the Sheraton Seattle Hotel. The address is: 1400 Sixth Avenue, Seattle, WA 98101.

Below are the floor maps of the Sheraton Seattle Hotel during the whole conference.

Second Floor

Grand Ballroom

A   B   C   D

Semantic Spaces
Temporal Data

Loading Dock

Aspen

Spruce

Willow A   Ed. Tools II
Willow B   Keynote

Cedar

Research Tools I
Research Tools II
Ed. Tools I

President's Address

MAP KEY

Restrooms

Stairs

Elevators

9:45 -11:00 - Session I
7:30 a.m. Registration out front of Willow A
8:00 - 9:30 Poster Session and coffee - (Location TBA)

1. Jamieson, Hanna, & Crump: A simple network model of associative learning (not memory)

2. Matthews, Jones, Rass, Hetrick, & O'Donnell: Semantic effects in letter fluency

3. Jones & Kievit-Kylar: The semantic pictionary project

4. Johns, Taler, Young, Sheppard & Jones: A computation analysis of semantic structure in bilingual verbal fluency tasks

5. Huang, Xie, Hu, Tang, & Starnes: Content coverage analysis based on DSSPP

6. Chubala & Jamieson: A holographic exemplar model of implicit learning

7. Gomez, Cox & Geller: Modeling corrective saccades

8. Whitson, Williams, & Harwood: A comparative study of visual-spatial performance with organic molecular representations using eye-tracking

9. Prince, Hawkins, Love, & Heathcote: An r package for state-trace analysis

10. Paxton & Dale: Multimodal synchrony: Tracking body and voice in an affordable behavioral recording setup

11. Liao & Sheu: Impacts of perceptual complexity on target detection performance: A nonlinear mixed model analysis

12. Ohyanagi, Sengoku, Miyazaki & Liu: A solution for measuring accurate reaction time to visual and auditory stimuli and its application for assessments in occupational therapy

13. Lawrence, Baugh & Marotta: Behavioural distinction between strategic control and spatial realignment during visuomotor adaptation in a viewing window task

14. van der Velden: Finger tracking in video using ridge detection

15. Stoet: Demonstrating cognitive experiments using PsyToolkit and Java

16. Plant, Quinlan, Reader & Thompson: DMTpsych: Postgraduate training for research data management in the psychological sciences

17. Hope, Destefano, Schoelles & Gray: Space fortress 5

18. Recchia, Kievit-Kylar, Jones & McRae: Using web games to elicit associative and feature-based conceptual representations

19. Christopher Koch: Considerations for online homework systems

20. Conley, Wutke, Calhoun & Evans: Video games and visual search: The roles of competition and experience

21. Hamm & Beasley: Balance beam aid for instruction in medical diagnosis

22. Wade, Taylor, Stancin, & Brown: Web-based problem-solving for teens with traumatic brain injury: Who benefits?

23. Brunstein, A., Brunstein, J., Pamulapati, Nour & Sargsyan: Computer versus faculty generated assessment of surgical simulation performance: What are the differences?

24. Ghiasinejad: Toward Automation of Verbal Protocol Data Analysis

25. Boonthum, Magliano, & McCarthy: Rethinking the automatic assessment of reading processes: Is it just a matter of doing more?

26. Chrea, Sheu, Wang, Navarro & Mougin: Automatic evaluations of advertising messages

Willow A		Cedar	
Session A: Semantic Spaces		**Session B: Educational Tools I**	
Chair: Carl Cai		*Chair: Christopher Wolfe*	
9:45	Cai, Graesser, Burkett, Li, Morgan & Shaffer	9:45	Wolfe, Hu, Reyna & Fisher
10:00	Zhao, Doyle-Smith & Li	10:00	Varner, Roscoe, & McNamara
10:15	Kievit-Kylar & Jones	10:15	Roscoe, Varner, Weston & McNamara
10:30	Johns & Jones	10:30	Taraban, Harold & Zeng
10:45	Burgess		

11:00 - 12:00 - Session II

Willow A		Cedar	
Session A: Temporal Data		**Session B: Research Tools I**	
Chair: Kenneth McRae		*Chair: Ping Li*	
11:00	McRae, Khalkhali & Wammes	11:00	Khan, Giles & Li
11:15	Plant & Turner	11:15	Sheu, Causeur, Chu & Hsieh
11:30	Peter Dixon	11:30	Armstrong, Watson & Plaut
11:45	Burgess, Chen & Dumon	11:45	Chu & Sheu

12:00 - 1:00 - Lunch break

1:00 - 3:00 - Presidential Symposium in Willow

Joe Magliano

Natural Language Processing Applications in Education: The Computer-based Analysis of Student Constructed Responses

3:15 - 4:15 - Session III

Willow A		Cedar	
Session A: Educational Tools II		**Session B: Research Tools II**	
Chair: Patricia Wallace		*Chair: Justin Estep*	
3:15	Wallace, Millis, Ide, Wilson, Graesser & Halpern	3:15	Estep & Burgess
3:30	Brandon, Roscoe, Tanner Jackson, Dempsey, & McNamara	3:30	Matsuka & Honda
3:45	Queen, Ray, R., Piselli & Ray, J.	3:45	Chen, Chung & Sheu
4:00	Oberjohn & Wade, Ph.D.	4:00	Cheng & Sheu

Keynote: 4:30-5:30 in Willow
Susan Dumais
Understanding in situ search behavior using large-scale log analysis and web-based experiments
Business meeting: 5:30-6:00

Keynote Speaker: Dr. Susan Dumais

Microsoft Research

Understanding in situ search behavior using large-scale log analysis and web-based experiments

 In recent years, the rise of cloud-based services has made it possible to capture human interaction with computing systems on a scale previously unimaginable. These large-scale logs represent traces of human behavior, seen through the lenses of the tools people use. Unlike other methods for observing behavior, logs provide a very rich picture of what people actually do in real-world task situations. Additionally, the scale at which log data can be collected gives a good picture of subtle behavior changes or unusual (so-called "tail") activities that could not possibly be observed by looking at only hundreds (or even thousands) of people. In addition to observing human behavior at scale, large-scale web services also enable controlled experimentation to evaluate alternative algorithms and interface designs. This talk will provide an overview of large-scale log analysis and experimental design, with examples from what we have learned from real world web search and browsing behaviors.

SCiP Presidental Symposium

Natural language processing applications in education: The computer-based analysis of student constructed responses

Joe Magliano

Northern Illinois University

Assessment of student knowledge and progress is an integral part of education for obvious reasons. We want to know if students are making expected progress as they progress through the educational system. However, assessment is also important for gauging student progress in the context of a curriculum or intervention so that appropriate feedback can be provided. Assessment often serves this important function in intelligent tutoring systems. This begs the question, in the context of these systems, what are the best types of behavior to assess? There is no simple answer to this question, but a number of researchers have opted to develop systems that rely on student constructed responses, such as short answers, summaries, essays, and typed verbal protocols (i.e., typed "think aloud" or self explanation verbal protocols). Using natural language algorithms to analyze constructed responses is a complex endeavor. There are a variety of approaches that can be employed that use of key word/phrase matching, semantic spaces (e.g., LSA, topic modeling), or a combination of the two. Which approach is best depends not only upon empirical evaluation, but an assessment of what one want to do with the assessments. In this Presidential Symposium, I will fist talk about the merits and challenges of using student-constructed responses in computer-based educational applications. The subsequent presentations are representative of some of the state-of-the-art applications of natural language processing in computer-based assessments and tutoring systems.

The growth of automated scoring of student responses with latent semantic analysis

Peter W. Foltz and Thomas K Landauer

Pearson's Knowledge Technologies

At SCiP in 1995, we presented a talk describing several approaches to using Latent Semantic Analysis for analyzing student written responses. Over the past sixteen years, the work has expanded both the theoretical basis of Latent Semantic Analysis as well as the types of applications to which it can be applied. This talk will address the aspects of LSA that allow the measurement of knowledge and writing abilities in constructed responses. The talk will further provide some examples of the range of constructed responses currently being scored operationally. These include assessing short answer questions, student summaries, essays, vocabulary growth and team communications.

Natural language processing to support large-scale formative assessments of verbal skill

Paul Deane, John Sabatini, Russell Almond, Bob Krovetz, Rene Lawless
Education Testing Service

Historically, systems that provide automated scoring for constructed verbal responses (e.g., writing) have been driven by the needs of summative assessment; but close integration between instruction and assessment requires a different approach emphasizing the interpretability of features. The kind of rethinking required can be illustrated by examining two dimensions of verbal (or writing) skill: word choice and transcription. *Word choice*. The ability to choose the right words can be measured using semantic space methods (LSA and its relatives), but at the cost of interpretability. We will discuss a corpus-based method for studying vocabulary and essay content, supervised topic maps, which links vocabulary usage patterns to human-labeled topics, and thus allows for an analysis in which patterns of word choice can be presented more transparently in an instructional context. *Transcription*. Success at the process of getting text on the page can be measured by length, elaboration of structure, or absence of mechanical and linguistic errors, but this measurement provides little useful information to instructors. Keystroke logs – detailed information about the timing of writing events – can be used to disambiguate the evidence provided by the final written product, and thus to distinguish students by the kind of writing performance they produce . These and other features can be targeted to provide formative feedback, as in an ITS, or to teachers, to provide actionable information about student performance profiles. But such goals can only be achieved by adapting NLP techniques to provide specific and, instructionally useful information about student performance.

Domain-specific semantic processing portal

Xiangen Hu, Jianmin Dai, & Douglas A. Starnes
University of Memphis

The Domain-Specific Semantic Processing Portal (DSSPP) is a collection of web services that provides semantic processing capabilities for researchers and developers. DSSPP is an implementation of a general framework of semantic spaces proposed by the Authors (Hu, 2005). DSSPP is capable of incorporating any vector-based semantic spaces. The current implementation is built upon a cloud computing platform that is fast and scalable. DSSPP is the technology that supports several applications developed in the ADL Center for Intelligent Tutoring Systems Research and Development (ADL CITSRD) at the University of Memphis. To highlight the usefulness of DSSPP, we will demonstrate 1) Learner's Characteristic Curves (LCC) that serves as student model in Intelligent Tutoring System, 2) three new semantic similarity metrics that measure semantic similarity between words from different semantic spaces, 3) a semantic spectrum analytic method that analyzes domain-specific semantic properties of text (words, sentences, conversations), and 4) a content coverage analysis tool that can be used as a part of essay grading.

Natural language processing in a writing strategy tutoring system

Danielle S. McNamara[1], Rod Roscoe[2], Scott Crossley[2], Zhiqiang Cai[2], and Art Graesser[2]

Arizona State University,[1] University of Memphis[2]

Historically, ITSs address upon well-defined, tractable domains such as algebra, vocabulary, or fact-based content (e.g., science and history). However, there is a growth ITSs that focus on ill-defined domains such as writing and strategy learning. These systems face the same challenges as ITSs in well- defined domains, but must overcome additional hurdles unique to ill-defined domains. The Writing Pal is such an ITS that provides high school and college students with training to use strategies to improve their writing. It includes lessons on the strategies, game-based strategy practice, and essay writing with feedback that focuses on the use of the strategies to improve the essays. Writing is an ill-defined domain because the features of "skilled" or "effective" writing are difficult to quantify, and individual writers may employ diverse strategies and styles to achieve similar goals. Moreover, the acquisition of these skills is an ill-defined, fluid process requiring extended practice with individualized feedback. The Writing Pal also departs from other traditional essay scoring systems because it does not focus on lower level trait feedback (grammar, mechanics, word usage), but rather centers the students' attention on the process of writing, and on strategies to improve that process. This presentation will describe success and challenges with the development of the computational linguistic algorithms needed to drive interpretation of students' responses during essay writing and practice. Unlike systems that provide holistic essay scores, these algorithms must be able to evaluate specific instances of strategy usage and inform specific, formative feedback on strategy improvement.

Assessing multiple source integration in student essays

Peter Hastings[1], Simon Houghes[1], Joseph P. Magliano[2], Susan R. Goldman[3], Kimberly Lawless[3]

DePaul University,[1] Northern Illinois University,[2] University of Illinois-Chicago[3]

A critical need for students in the digital age is to learn how to gather, analyze, evaluate, and synthesize complex and sometimes contradictory information across multiple sources and contexts. In a recent study, we gave 459 Chicago middle school students three short articles that described complementary factors leading to the growth of Chicago in the mid-1800s. Then the students were asked to write their own essay on the topic, "Why Chicago became a big city." The essays were annotated to indicate, on a sentence-by-sentence basis, how they corresponded to two different representations of the original source texts: the text model (original source sentences) and documents model (graphical semantic representation). With this corpus, we evaluated three different automatic text processing mechanisms for their abilities to replicate the human analysis. The first was a simple pattern-matcher that allowed flexible matching of words and phrases in the sentences. This was supported by a web-based tool that allowed a user to quickly enter patterns and see almost immediately their performance on the corpus. The second technique was Latent Semantic Analysis, which was used to compare student sentences to original source sentences using its high-dimensional vector-based representation. We also used a popular Machine Learning technique, Support Vector Machines, to learn a classification scheme from the corpus. In this talk, I will describe the different

techniques in more detail, and discuss their strengths and weaknesses for automatically inferring how well the students integrated material from the different sources.

Natural language assessment within game-based practice
Tanner Jackson[1] & Danielle S. McNamara[2]
University of Memphis[1], Arizona State University[2]

Intelligent Tutoring Systems (ITSs) are situated in a constant struggle between effective pedagogy and system enjoyment and engagement. Natural Language Processing (NLP) has been an integral component to many ITSs, and it allows learners to use their own words and ideas. Unfortunately this flexibility isn't always enough to fully engage students, and recently ITS researchers have turned to games and game-based features to help increase engagement and enjoyment. Combining NLP and games should provide both the benefits of personalized instruction from NLP with the added enjoyment and persistence associated with games.

9:45-10:45 Parallel Session I

Session A: Semantic Spaces

9:45 **Generation of domain-specific LSA spaces**
Zhiqiang Cai, Arthur C. Graesser, Candice M. Burkett, Haiying Li, Brent Morgan, University of Memphis
David W. Shaffer, University of Wisconsin
zhiqiang.cai@gmail.com, art.graesser@gmail.com, candicemburkett@gmail.com, haiyinglit@gmail.com, bmorgan2@memphis.edu, dws@ef-games.com

The semantic representation of words in LSA spaces is constrained by the corpus used to generate the space. While generating a space is not as challenging as it used to be, identifying a domain-specific corpus remains a non-trivial issue. To address this issue, we developed an algorithm to automatically identify a domain-specific corpus on a specific subject matter and subsequently generate a domain-specific LSA space. The algorithm requires (a) a small domain-specific corpus as a "seed" corpus, (b) a large universal corpus that contains documents about all domains (e.g. Wikipedia), and (c) a general-language corpus that reflects typical language in print (e.g. TASA). The algorithm first extracts words from the seed corpus and computes "keyness" of each word based on its frequency in the seed corpus and inverse entropy weight in a reference corpus (e.g., TASA). Documents are then selected from the universal corpus based on the total sum of keyness in the documents. The selected documents are then merged into the language supporting corpus to generate the space. Using this algorithm, we created an LSA space on urban science for a learning environment we are developing (i.e., an epistemic game with an automated mentor). The small seed corpus consists of 14 articles about urban science. Wikipedia was used as the universal corpus and TASA as the everyday language corpus. The computed keyness correlated 0.70

13

(N=90) and the LSA similarity correlated 0.62 (N=80) with human's ratings. This algorithm greatly reduced the amount of work required in collecting the domain-specific corpus.

10:00 **A comparative study of semantic representations across three languages**

Xiaowei Zhao, Noah Curran Doyle-Smith, Emmanuel College
Ping Li, Pennsylavania State University
zhaox@emmanuel.edu, doylesmithn@emmanuel.edu, pul8@psu.edu

Zhao, Li and Kohonen (2011) recently introduced a software package (Contextual Self-Organizing Map) that applies a corpus-based statistical learning algorithm to derive semantic representations of words. The algorithm relies on the analyses of contextual information extracted from a text corpus, specifically, analyses of word co-occurrences in a large-scale electronic database of text. In this current study, we applied the program to extract semantic representations of words in three different languages: English, Spanish and Chinese. In order to conduct a more accurate comparison across the three languages, we used the translations of the same material as the basic corpora 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 study. For each language, the derived semantic representations of the most frequent 300 words in the corpus were further processed by a self-organizing map, an unsupervised neural network model that projects the statistical structure of the context onto a 2-D space. Our results demonstrate the ability of statistical learning based on word co-occurrence to capture the semantic relationships of words across typologically different languages. For all three languages, words with similar meanings cluster together, forming groups that correspond to lexically meaningful categories on self-organizing maps. In addition, important or unique linguistic characteristics of each language are also captured in the derived semantic representations. For example, words with the same gender in Spanish often group together on our derived map, as are classifiers in Chinese and inflected past tenses or past participles in English. We have also conducted a comparison of our method with some other methods like HAL (Burgess & Lund, 1997) and BEAGLE (Jones & Mewhort, 2007) and will report results of this comparison at the SCiP meeting.

10:15 **Word2Word: A visualization tool for high-dimensional semantic data**

Brent Kievit-Kylar, Michael Jones
Indiana University

Although advances in statistical semantic tools are made every year, the techniques used to evaluate these systems have advanced little. Typically, the plausibility of a semantic space is explored by sampling nearest neighbors to a target and evaluated based on the modeler's intuition. Tools for visualization of these spaces have been sorely lacking. We present a new open-source tool to plot and visualize semantic spaces, allowing researchers to rapidly explore patterns in visual data representative of statistical relations between words. Words are visualized as nodes and word similarities as directed edges of varying strengths.

The "Word to Word" visualization environment allows easy manipulation of graph data to test word similarity measures on their own accord or as a comparison between multiple similarity metrics. We also provide a large library of ready to use, modern, statistical relation tools along with an interface to teach them from various language sources. The modularity of the visualization environment allows for quick insertion of new similarity measures to compare new work against current state of the art.

10:30 **Grounding the lexicon: Inferring perceptual representations from global lexical similarity**
Brendan T. Johns, Michael N. Jones
Indiana University, Bloomington
johns4@indiana.edu, jonesmn@indiana.edu

Distributional models of semantic memory learn the meaning of words through exposure to the linguistic environment. Although these models have been demonstrated to account for a large and diverse set of data (see Riordan & Jones, 2011 for a review), it has become clear that perceptual information must be integrated into distributional models to best explain human data. A variety of results, across both behavioral and neuroimaging experiments, has shown the necessity of perceptual information in language processing (see Barsalou, 2008 for a review). One important step that has been taken to make this possible is the collection of feature norms across a large number of words (e.g. McRae, et al., 2005). However, these norms only encompass approximately 500 words. In order to integrate these types of representation into large-scale models of semantic memory, it would be ideal to be able to infer feature representations for the entire lexicon. We propose a method to estimate a word's feature representation based on the associative structure of words in a large corpus. Specifically, the method uses the global structure of a distributional lexicon to exploit the redundancy between language and perception in order to generate perceptual representations for words with which the model has no perceptual experience. We will demonstrate how these inferred representations can be used to simulate data based on word-word relationships, and also how they can be integrated into a new sentence-processing model to explain more sophisticated data.

Session B: Educational Tools I

9:45 **Teaching conditional probability estimation with AutoTutor Lite: Semantic coherence and breast cancer risk**
Christopher R. Wolfe, Miami University
Xiangen Hu, University of Memphis
Valerie F. Reyna, Cornell University
Christopher R. Fisher, Miami University
WolfeCR@muohio.edu, xiangenhu@gmail.com, vr53@cornell.edu, fisherc2@muohio.edu

We used AutoTutor Lite to create a prototype tutorial to teach conditional probability estimation. AutoTutor Lite lets people interact with a talking animated agent by entering responses in a textbox. It answers questions, provides feedback,

asks for information, gives hints, and summarizes responses. Graphical displays include animation or video. AutoTutor Lite works on many platforms using an ordinary browser. It provides a number of helpful authoring tools. We tested participants' ability to make conditional probability judgments pertaining to the relationship between breast cancer and genetic risk – the presence of BRCA1/2 genetic mutations. The quality of conditional probability estimates was assessed in three ways. First, in terms of semantic coherence, an internal consistency benchmark; second as the degree of internal inconsistency; finally as empirical accuracy compared to statistical norms. Two experiments compared versions of individual face-to-face human tutoring to a no-intervention control, and AutoTutor Lite to two comparison groups. In Experiment 1, as predicted by Fuzzy-Trace Theory, individual human tutoring was superior to the control group whether teaching the logic of the 2x2 Table (best), using Euler diagrams, or using relative frequencies. In Experiment 2, AutoTutor Lite was compared to a text-only tutorial and a control group. The AutoTutor Lite group performed slightly, but not significantly, better than the other groups. The performance of the human tutored groups in Experiment 1 was substantially superior. AutoTutor Lite provides a powerful set of tools that can be utilized effectively. However, individual human tutoring remains the "gold standard" for future work in this area.

10:00 **Linguistic alignment of student and teacher holistic ratings of essay quality**
Laura K. Varner, Rod D. Roscoe, Danielle S. McNamara
University of Memphis, Arizona State University
Laura.varner22@gmail.com

National assessments demonstrate that many students are writing below proficiency (National Assessment of Educational Progress, 2002; National Commission on Writing, 2008). One problem may be the misalignment of students' and teachers' criteria for "good writing." Human detection of these misalignments is inconsistent due to the subjective and complex nature of writing. However, automatic computer software may be able to provide a more objective analysis of these semantic misalignments. This detection could reveal opportunities for revised pedagogy and remediation. In this study, we used the Linguistic Inquiry and Word Count database (LIWC, 2007), a software tool for semantic text analysis, to explore student and teacher ratings of essay quality. We collected 126 timed, prompt-based, persuasive essays from high school students near Washington DC. Students were asked to predict their score on a 1-6 scale. Teachers rated each essay on a 1-6 scale using the College Board SAT rubric. Pearson correlations were calculated comparing student and teacher scores to LIWC word categories. The analysis suggests that students tended to associate essay quality with confidence words, strong vocabulary, and the objectivity of their examples. Teachers similarly assessed essay quality based on confidence words, strong vocabulary, and objective language. However, teachers' assessments were also positively related to idea elaboration and negatively related

to hypothetical language, possibly suggesting teachers were more sensitive to the appropriate development of persuasive evidence. Students tended to perceive the strength of their arguments and ideas, but lacked a complete understanding of the proper explanation and defense of such ideas.

10:15 **Mixed-method usability testing and development of the writing pal intelligent tutor**
Rod D. Roscoe, Laura K. Varner, Jennifer L. Weston, Danielle S. McNamara
University of Memphis, Arizona State University

Writing Pal (W-Pal) is an intelligent tutor (ITS) that offers writing strategies, game- and essay-based practice, and formative feedback for high school and college students (McNamara et al., 2011). These pedagogical principles are critical to writing proficiency development (Graham & Perin, 2007). However, designing a complex ITS in an ill-defined domain presents numerous challenges, such as how to communicate open-ended strategies and feedback while maintaining engagement. Typically, ITS usability testing is an iterative process wherein a few students (i.e., end-users) interact with specific system components and provide feedback. The system is modified and the process repeats until recommended changes are minimal. In this paper, we discuss how W-Pal development has benefitted from a mixed-method usability process combining traditional iterative usability with two alternative formats. In *in-house usability*, developers adopt the student role to use the system following explicit user constraints. This method violates the principle of soliciting end-user feedback, but is faster and still highly informative (Roscoe et al., 2011). Moreover, it can reveal pedagogical issues *prior to* deployment with students, which is an important ethical consideration. In *in vivo usability*, a complete system is deployed in an authentic context. In this case, 140 10th-graders used W-Pal for six months in their English classrooms. This method violates the principle of rapid iterative testing, but allows users to experience and provide feedback upon the entire system in its intended context. Importantly, some usability issues for complex ITSs in ill-defined domains can *only* emerge in the context of longitudinal, *in vivo* testing.

10:30 **Measures of quality in research papers**
Roman Taraban, Stephanie Harohd, Xiaofang Zeng
Texas Tech University
roman.taraban@ttu.edu, stephanie.harold@ttu.edu, xi.zeng@ttu.edu

The development of expository writing is universally regarded as important for undergraduates in scientific disciplines, yet remains a challenge for both students and faculty. One limiting factor in students' development is the limited opportunities they have to actually write and receive constructive feedback. In part, machine solutions have been pursued in an attempt to expand student access to formative learning opportunities. This includes applications of latent semantic analysis (LSA) and automated systems that track multiple dimensions of text composition, like Coh-Metrix. The present work is not a machine implementation

but rather an examination of discourse and syntactic properties of research papers. The goal is to determine how much of the variance in a human-assigned grade can be accounted for by these properties as a precursor to consideration of machine applications. In this study, we analyzed archived research papers written by undergraduates who had participated in intense mentored research experiences during the academic year (approx. 10-15 h / week in a lab). Participants were typically seniors (mean completed credits = 107; range 61-137). One "gold standard" against which to judge these papers was an identical analysis of journal articles by students' mentors and other faculty. A second "gold standard" was standard grades (e.g., A, B, C) assigned to these papers by instructors. Fifty archived research papers (typically about 10 pages, with figures, tables, and references) submitted by students at the end of the academic year, and six journal articles written by faculty on topics similar to students' topics were analyzed. Two kinds of analyses were carried out: analyses of discourse complexity and analyses of syntactic complexity. Analyses were conducted at the sentence level. To carry out the discourse analyses, a taxonomy of kinds of discourse functions and a rubric were created through the consensus of three experimenters. The syntactic analyses consisted of two measures: length of *t-units* (essentially, sentence length) and frequency of nominal clauses. Faculty papers exceeded student papers in the use of higher-order discourse functions and syntactically complex constituents. In this sense, the rubrics for discourse and syntactic analyses enable the identification of upper benchmarks for student compositions. Discourse functions, t-unit length, and number of nominal clauses correlated with instructor-assigned grades and accounted for significant variance. Overall, the discourse and syntactic elements were validated through these analyses as related to the quality of student expository composition. They are well defined, and the syntactic elements, especially, lend themselves to machine analysis. Limitations of these measures are also considered.

11:00-12:00 Parallel Session II

Session A: Temporal Data

11:00 **Integrating words that refer to temporally sequenced events**
Ken McRae, Saman Khalkhali, Jeffrey Wammes
University of Western Ontario
kenm@uwo.ca, saman.khalkhali@gmail.com, jwammes2@uwo.ca

The organization and processing of event concepts in semantic memory is an important issue in language processing and memory research. We tested whether pairs of words denoting events that plausibly occur in sequence (*marinate-grill*) generate expectancies for a target that denotes a subsequently occurring event (*chew*). Experiment 1 showed that two events in sequence prime the third. Experiment 2 showed that the individual primes (i.e., *marinate* and *grill* separately) do not prime their related event targets. Therefore, information from both primes must be integrated to sufficiently activate knowledge of the

subsequently occurring target. This is the first study to demonstrate priming among words denoting sequentially occurring events. In Experiment 3, a relatedness decision task, the processing of the event triplets was facilitated when the first two event words were presented in a temporally correct order (*marinate-grill-chew*), as compared to when their order was reversed (*grillmarinate-chew*). These results cannot be accounted for by spreading activation theory, at least in terms of associative connections measured by the strength of normative association, because associations among the words in the triplets were extremely low. The results also pose problems for LSA because it is not sensitive to word order. They may be accounted for by models in which word order or differential conditional probabilities are encoded, as in BEAGLE. Finally, they provide evidence for the role of situation models and the use of world knowledge during online language comprehension, even in the absence of sentential contexts.

11:15 **Do I really need to worry about millisecond timing any more?**
Richard R. Plant, Garry Turner
University of York, UK

Psychologists are, and have always been, concerned about millisecond timing accuracy if one traces back to the development of the tachistoscope by Wundt (c.1875). In many areas contemporary researchers regularly make use of commercial and custom written software to administer paradigms. Increasingly they use complex multimodal stimuli interacting with hardware which would have been unimaginable a few years ago. It is common to see studies where conditional differences are in the order of tens of milliseconds. Some less. We discuss whether some modern equipment is actually worse than that of 136 years ago? Has experimental rigour worsened in the last decade even? Can anything be done from a practical standpoint? Illustrations from various research areas will be discussed. In common with other scientific disciplines we would like to see researchers quoting they have actively checked their equipment and foresee self-validation as a way forward for today's complex computer-centric studies.

11:30 **Mixture or shift? A simple diagnostic for response time distributions**
Peter Dixon
University of Alberta
peter.dixon@ualberta.ca

Response time distributions are generally mixtures of different levels of preparation,strategy, and criteria. Nevertheless, in our analyses of response time data, we typically assume that experimental manipulations simply shift the distribution rather than affect the nature of that mixture. In the present paper, I present a simple, distribution-free technique for deciding whether an experimental manipulation is best thought of as shifting distributions or mixing distributions. In an experiment that involves three ordered conditions, one first estimates quantiles for conditions A (the fastest) and C (the slowest). These quantiles are used to derive the best-fitting quantiles for the intermediate condition, B, under the

assumption either that condition B is a mixture of A and C or that B is a shift between A and C. The log likelihood of the data, given the estimated quantiles, provides a measure of the evidence in favor of a mixture relative to a shift. Monte Carlo simulations demonstrate that this approach is efficient and robust with even a modest amount of data. The technique is applied to data from a task-switching paradigm (in which theory predicts a distribution mixture) and a psychological refractory period paradigm (in which theory predicts a shift). I also present some surprising results from other paradigms in which there are no obvious theoretical prediction.

11:45 **Chronological shift in the connotation of the meaning of recovered memory**
Curt Burgess, Christopher L. Chen, Victoria N. Dumon
University of California, Riverside
curt@ucr.edu, cchen033@ucr.edu, vdumo001@ucr.edu

The conceptualization of recovered (or repressed) memory has changed as empirical and legal evidence has accumulated over the last three decades. Claims of repressed memories became admissible in court in the 1980's. The 1990's was a decade of increasing legal challenges to cases where individuals had been found guilty of sexual abuse based on such testimony, followed by many of these cases being overturned. In this study we conducted a semantic analysis of the newsletters of the partisan False Memory Syndrome Foundation that ranged from March 1992 to Spring 2011. This corpus totaled 2718354 words. Two high-dimensional semantic models were constructed using the HAL model. The early model used the earliest newsletters (the first third of the entire corpus) whereas the later model used the last third of the corpus. The models were used to generate semantic neighborhoods and distances between word pairs. Raw word frequency was informative as well. First we determined that the models had sufficient input text to form meaningful representations. This was accomplished by noting that neighborhoods of family members (mother, father, etc) contained other family and related words as one example. We also found basic relatedness effects (related distances < unrelated distances). A subset of the results include a substantial drop from neighborhoods of MPD. The notion of "repression" is less contextually related (compared to "recovered") in the later model. Results include other concepts "accusation" vs "allegation;" "incest," memory "validity." We hope to have an additional comparison by conference time that would include a text sample from a contrary partisan group.

Session B: Research Tools I

11:00 **BrainSeer: A neuroinformatic tool for functional brain imaging based on automatic information**

Omair A. Khan, Lee Giles, Ping Li
The Pennsylvania State University
oak3@psu.edu, giles@ist.psu.edu, pul8@psu.edu

BrainSeer is a novel neuroinformatic tool that will allow researchers to have easy, customized access to fMRI data. It aims at generating a database containing brain areas that become active in response to specific experimental tasks, along with parameters of scanning, demographics, etc., in published fMRI studies. Currently available applications in this domain are limited in that 1) information cannot be easily searched, 2) information display is solely based on anatomical regions, and 3) they are unable to derive metaanalyses on the fly. BrainSeer addresses these issues by providing researchers with multiple options for various purposes. In particular, it has the ability to extract and concatenate data from tables in PDF files. Users can customize searches with operators such as anatomical regions of interest (ROI), experimental parameters, table captions, or specific participant characteristics. In addition, BrainSeer provides users with an interactive interface of clickable brain regions, pinpointing to specific ROIs that can then be used as search input. Researchers can see details of the fMRI studies that show brain activities in or around the ROI for further search refinements. In sum, BrainSeer is webbased, user-friendly interface through which investigators can identify the brain as coordinated neural networks, and as such it will greatly enhance the cyber-infrastructure for cognitive neuroscience and neuroinformatics.

11:15 **False discovery rate and factor analysis for multiple testing in ERP data analysis**

Ching-Fan Sheu, National Cheng Kung University, Taiwan
David Causeur, Agrocampus Ouest, France
Mei-Chen Chu, Shulan Hsieh, Cheng Kung University, Taiwan
csheu@mail.ncku.edu.tw, David.Causeur@agrocampus-ouest.fr,
clap47@msn.com, psyhsl@mail.ncku.edu.tw

Event-related potentials (ERP) are widely used by psychophysiological researchers to determine the time courses of mental events. When comparing event-related potentials across different experimental conditions, often there is no a priori information on when or how long the differences should occur. Testing simultaneously for differences over the entire digitized time intervals encounters serious multiple comparison problem in which the size of false positive error must be kept low, while attaining reasonable power for correct detection. This study compared the performance (on real and simulated data sets) of Bejamini-Hochberg (1995) false discovery rate procedure (BH) and a recently developed factor-adjusted multiple testing procedure (FAMT) under dependence (Friguet, Kloareg, & Causeur, 2009) in managing the multiple comparison problem in ERP

data analysis. The results showed that, in most cases, the FAMT procedure outperformed the BH procedure in detecting the number of signiffcant intervals, suggesting that the BH procedure might over-control the false discovery rate under strong dependence data structure such as ERP measurements.

11:30 **SOS! An algorithm and software for the stochastic optimization of stimuli**
Blair C. Armstrong, Carnegie Mellon University
Christine E. Watson, University of Pennsylvania
David C. Plaut, Carnegie Mellon University
blairarm@andrew.cmu.edu, watsonc@mail.med.upenn.edu, plaut@cmu.edu

The characteristics of stimuli used in an experiment critically determine the theoretical questions it can address. Yet, there is relatively little methodological sophistication in how researchers select optimal sets of items, and most researchers still carry out this process by hand. In this research, we present SOS, an algorithm and software package for the Stochastic Optimization of Stimuli. SOS takes its inspiration from a simple manual stimulus selection heuristic that has been formalized and refined as a stochastic relaxation search. This algorithm rapidly and reliably finds stimuli that optimally satisfy the constraints imposed by an experimenter. This allows the experimenter to focus on selecting an optimization problem that suits their theoretical question and not on the tedious task of manually selecting stimuli. We detail how this optimization algorithm, combined with the vocabulary of constraints that define optimal sets, allows for the quick and rigorous assessment and maximization of the internal and external validity of experimental items. In doing so, the algorithm opens up new avenues for conducting research using classic (e.g., factorial) and more contemporary (e.g., mixed-effects regression) experimental designs. We demonstrate the ease of using SOS with a case study and discuss other research situations that could benefit from this tool. Support for the generality of the SOS algorithm is demonstrated through Monte Carlo simulations on a range of optimization problems faced by psychologists. The software implementation of SOS and a user manual are provided free-of-charge for academic purposes as pre-compiled binaries and as MATLAB source files at: http://sos.cnbc.cmu.edu.

11:45 **Quantifying age differences in the affective evaluation of chinese antonyms**
Mei-Chen Chu, Ching-Fan Sheu
National Cheng Kung University, Taiwan
Clap47@msn.com, csheu@mail.ncku.edu.tw

Studies on human emotion often require participants to evaluate affective stimuli such as words or pictures. Typically, participants choose one response category from several arranged in hierarchical order. Often it is desirable to construct a reference ordering for a set of stimuli (a norm) against which reactions toward these items might be compared between individuals or groups of individuals. A number of shortcomings might be identified for affective norms in current use. Most norms use mean ratings to order items even though mean scores may not be

appropriate summaries for discrete levels of responses. Most norms are also designed to gauge participants affective reaction to the same item in more than one dimension. Thus, how ratings from different dimensions should be integrated becomes an issue. In addition, few word norms include considerations of how item responses may be affected by population characteristics of respondents such as age, although many life-span studies reported that younger and older adults tend to judge stimulus valence differently. This study investigated older and younger adults reactions to a set of Chinese antonyms in order to quantify age effect within the framework of a multidimensional item response theory (IRT). A multidimensional IRT model was proposed to calibrate participants responses from four different affective dimensions. Our results provided support for the aged-related positivity effect usually found in aging research and suggested that future development of affective norms should adjust for age.

3:15-4:15 Parallel Session III

Session A: Educational Tools II

3:15   **Using AutoTutor in a serious game**
Patty Wallace, Keith Millis, Michelle Ide, Susan Wilson, Northern Illinois University
Art Graesser, University of Memphis
Dian Halpern, Claremont McKenna College
pwallace@niu.edu, kmillis@niu.edu, mide16@niu.edu, positivepwr@yahoo.com, art.graesser@gmail.com, diane.halpern@cmc.edu

AutoTutor is an intelligent tutoring program that teaches through interactive dialogs using natural language. Prior research has shown that AutoTutor increases learning in a variety of domains, including physics and computer literacy. To date, however, AutoTutor has not been used in gaming contexts. It is unknown whether AutoTutor would be successful in a gaming context because of gaming attributes (e.g., points, story line, etc) and player emotions (e.g., happy mood) that might distract from content learning. This paper describes and tests a portion of a serious game called "Operation ARIES!" which uses AutotTutor as its NLP engine. The goal of the game is to teach players to critically evaluate research reports found in various media. The game is meant for college freshman and high school seniors. The study reported here examined learning gains from one of the game's components: Case Studies. In this module, the player reads and evaluates several research descriptions. The basis of evaluation is whether the case contains one or more of 12 flaws, such as a lack of a control group, or an invalid dependent variable. In this module, the human player competes against a virtual student for points in the midst of an alien-themed story. In one study, we found greater learning gains in the game as compared to a non-game control group. In another study, we tested whether the presence of a storyline affected learning. We found that the story had little impact on learning. Therefore, AutoTutor appears to be useful in gaming contexts in addition to intelligent tutors.

3:30 **Persistence, performance, and perception: An analysis games in a writing strategy ITS**

Russell D. Brandon, Rod D. Roscoe, G. Tanner Jackson, Arizona State University
Kyle B. Dempsey, The University of Memphis
Danielle S. McNamara, Arizona State University
russell.brandon87@gmail.com, work.roscoe@gmail.com, gtannerjackson@gmail.com, kyle.b.dempsey@gmail.com, dsmcnamara1@gmail.com

The acquisition and mastery of writing skills and strategies typically require extended and deliberate practice (Graham & Perin, 2007; Kellogg & Raulerson, 2007). Such practice can induce fatigue and boredom, which hinders the learning process, but game-based methods of learning may help to ameliorate these effects (Gee, 2008; Van Eck, 2006). Games are hypothesized to support higher levels of motivation and enjoyment that may encourage continued practice. Writing Pal (W-Pal) is an intelligent tutoring system designed to teach writing strategies to high school and early college students. The system incorporates a suite of mini-games that support diverse modes of practice for specific writing strategies (Dai, Raine, Roscoe, & McNamara, 2011), such as identification, application, and generation. Decisions regarding game mechanics and aesthetics were constrained and informed by these different forms of practice (Hunicke, LeBlanc, & Zubek, 2004). Similarly, a variety of game elements (e.g., points, levels, fantasy, and narrative) were used to enhance motivation to persist in the practice. In this study, 140 tenth-grade English students used Writing Pal over one academic school year. Student perceptions of the games were collected via survey data, and persistence and performance data were obtained based on log-files collected as students played each game. This study examines different game features, and how these elements affected student perceptions of the system, persistence, and performance. For example, students overall rated the games as somewhat helpful and enjoyable. However, games that were rated more highly on these measures were played more often and for longer periods.

3:45 **The AVCA: An instrument for measuring audio visual comprehesion in college students**

Jennifer S. Queen, Roger D. Ray, Katherine Piselli, Rollins College
Jessica M. Ray, University of Central Florida
jqueen@rollins.edu, rdray@rollins.edu, kpiselli@rollins.edu, jessica.m.ray@gmail.com

This presentation introduces an instrument for assessing general Audio-Visual (AV) comprehension skills. Multimedia is ubiquitous in today's internet culture and is widely used in education as an alternative to traditional text-based supplements to instruction (Mayer, 2002). Unfortunately AV materials are not often used effectively, especially in the classroom (Hobbs, 2006). Likewise, little concern with the assessment of student readiness to learn from AV materials is

evident. Standardized tests exist for assessing reading comprehension at all educational levels, including SAT and GRE testing for college or graduate school readiness. But virtually no equivalents assessing AV comprehension exist, even though traditional college lecture classes depend heavily upon such presentations. Thus a standardized Audio Visual Comprehension Assessment (AVCA) instrument is needed for measuring the efficacy of AV materials and comprehension development. Our AVCA is a dual-form instrument for assessing student comprehension based on content in two independently presented "Virtual Laboratory Tour" (VLT) video productions. Each VLT presents a laboratory experiment and includes an introduction to the research problem, a detailed illustration of method, the results obtained, and a discussion of implications. An initial evaluation study involving 41 college student participants was conducted using a pool of 60 questions covering each video administered in a classroom setting. A subsequent large-scale validation study based on the 30 most reliable questions from each form is the next phase in progress. For this project we have converted our initial paper-and-pencil test to an internet-delivered 30-question-per-form electronic version that will be demonstrated in our presentation.

4:00   **Using technology to teach and coach positive parenting skills**
Karen S. Oberjohn, Shari L. Wade
Cincinnati Children's Hospital Medical Center
Karen.oberjohn@cchmc.org, Shari.Wade@cchmc.org

Training in positive parenting skills has been shown to significantly improve both parent and child behaviors. The most effective parenting skills programs combine didactic information with live coaching of the parent interacting with the child. Most of these programs traditionally require families to attend several (10 to 20) weekly sessions at behavioral health clinics. However, many families who need parenting skills coaching do not receive it due to barriers such as time, distance, transportation, sibling child care, and the unavailability of skilled providers. Inexpensive off-the-shelf technology makes it possible to overcome these barriers by delivering parenting skills training via standard high speed internet. I-InTERACT (Internet-based Interacting Everyday: Recovery After Childhood TBI) is an innovative 7-17 session web-based parenting skills program for families of children with TBI that combines self-guided web modules with live coaching via videoconferencing. Each web module provides a "theme for the week" (for example "Stress Management", "Dealing with Behavior in Public", "Cognitive Changes") with parent testimonials, parenting "how-to" videos, and didactics. Following completion of each web module families videoconference (using Skype or Cisco Telepresence) with a therapist to review the week's web module and play with their child while receiving bug-in-the-ear coaching (using a standard WebCam and BlueTooth or Microsoft LifeChat wireless headsets. Observational and self-report data suggest that this delivery method is feasible, effective at changing parent behavior, and very well received by parents, children and therapists alike.

Session B: Research Tools II

3:15 **Using r to harvest and preprocess text**
Justin Estep, Curt Burgess
University of California, Riverside
jeste002@ucr.edu, curt@ucr.edu

Web sites are an obvious candidate for sources of large amounts of text that is increasingly needed by researchers using statistical learning procedures and computational modeling. It is these cases in which labor intensive manual text extraction must be avoided. R is a multiuse program that is becoming increasingly popular for conducting statistical analyses but can serve many other functions as well. This talk will discuss the potential of R for harvesting large quantities of text from online sources as well as the preprocessing of text by the use of a series of rules and filters in the R program. Our labs use of the program for capturing individual blog information from a large social networking site will be used as a case study. The focus of the talk is exposure to R as an invaluable resource for working with text but will also cover usage of R, tips, benefits, and limitations.

3:30 **Identifying a set of effective models in factor analysis with a multiobjective optimization technique**
Toshihiko Matsuka, Hidehito Honda
Chiba University
matsukat@muscat.L.chiba-u.ac.jp, hito@muscat.L.chiba-u.ac.jp

Factor analysis is a statistical technique that is primarily used to describe structural relationships among variables. There is a technique called exploratory factor analysis (EFA), but it is exploratory in a very limited sense. In EFA, it is usually assumed that all manifested variables are associated with all latent variables, and then the strengths of associations are used to infer that structural relationships. In the present research we introduce a thorough model exploration technique for factor analysis using multiobjective stochastic optimization in order to identify a set of effective models. Van der Mass, Raijmakers, and Visser (2005) proposed a similar technique, yet their method was able to identify the "best" model with respect to the fit index for a given data set. In our approach, models are optimized for both complexities and fits to identify the Pareto-optimal models, i.e., the best fit model for a given level of model complexity or the least complex model for a given level of model fit, allowing a thorough exploration of model space and providing a diverse set of effective models. Examples are provided to show that examining a diverse set of effective models can improve understanding the structural relationships among variables and/or facilitate development of new hypotheses or theories about the relationships.

3:45   **Visualizing multilevel data**

Shu-Ping Chen, National Chengchi University
Yi-Hsiu Chung, National Chung Cheng University
Ching-Fan Sheu, National Chenk Kung University
Abby.chen.psy@gmail.com, royobarisk@pchome.com.tw, csheu@mail.ncku.edu.tw

Multilevel models have been widely used for the analysis of clustered or longitudinal data. The complex nature of these types of data appears to have led researchers to ignore the importance of exploratory data analysis in favor of examining solely numerical output for inferential purposes. In this work, we show that many fundamental visual-graphical methods can be used to help researchers better understand both multilevel data and models. In particular, we demonstrate that graphical techniques developed for meta-analysis can also be fruitfully employed in multilevel analysis. We implement these graphical routines in the open-source software R and illustrate them with two real data examples.

4:00   **Fitting hierarchical finite mixture models to response time data from the task switching paradigm**

Chung-Ping Cheng, Ching-Fan Sheu
National Cheng Kung University, Taiwan
cpcheng@mail.ncku.edu.tw, csheu@mail.ncku.edu.tw

Executive control is often studied by the task switching paradigm. Mean reaction time differences between switched and non-switched conditions are indications of switch cost. The failure-to-engage theory (FTE) assumes two components, prepared and unprepared processes, to account for the response times of a switching task. It explains the time differences between conditions by different mixing proportions. Standard response time analysis based on the FTE is performed in two stages. Reaction times of a participant given a condition are analyzed assuming a two-component mixture distribution and the extracted parameter estimates are then tested for differences between experimental conditions. This approach is inefficient because standard errors of parameter estimates obtained in the first stage are not accounted for in the second stage of the analysis. In this work we propose a single statistical framework for the analysis of reaction time data arising from the task switching paradigm. A hierarchical finite mixture model is used to fit simultaneously reaction times of all participants for all conditions. Various distributions such as Gaussian, ex-Gaussian and ex-Wald are considered for response times. We illustrate our approach with real and simulated data examples.

8:00-9:30 Poster Session

1. **A simple network model of associative learning (not memory)**
Randall K. Jamieson, University of Manitoba
Samuel D. Hannah, University of Queensland
Matthew J. C. Crump, Brooklyn College
randy_jamieson@umanitoba.ca, hannah.sam@gmail.com,
mcrump@brooklyn.cuny.edu

   We show that a simple neural network model coupled with a distributed stimulus representation scheme (Ghirlanda, 2005) reproduces a broad array of associative learning effects. Although the model's fit to data confirms it as a valuable tool, we argue that it nevertheless suffers an epistemic shortcoming. Namely, the model learns without remembering. We compare the simple network model to a recent exemplar model of associative learning based in Hintzman's (1986) Minerva 2 model for human memory (Jamieson, Crump, & Hannah, 2011). We use the contrast to underscore several meaningful distinctions between network and exemplar models of learning.

2. **Semantic effects in letter fluency**
Sean C. Matthews, Michael N. Jones, Olga Rass, William P. Hetrick, Brian F. O'Donnell
Indiana University
seamatth@indiana.edu, jonesmn@indiana.edu, rasso@indiana.edu,
whetrick@indiana.edu, bodennel@indiana.edu

   Verbal fluency tasks are commonly used to assess the integrity of the semantic memory system, and have been used to argue for the existence of semantic memory deficits in individuals with schizophrenia. It is typically assumed that letter fluency and semantic fluency use very similar search and retrieval mechanisms, but that letter fluency operates over phonological representations rather than the semantic representations targeted by semantic fluency. While there is some evidence that semantic factors may also play a role in the letter fluency task, the statistical methods that have been used to test this hypothesis rely on multiple participants producing the same items. This artificially restricts the analysis to only high-frequency words, and therefore cannot detect any semantic effects that may be present in the rest of the letter fluency data. In order to avoid this issue, we used a computational model of lexical semantics (Jones & Mewhort, 2007) to calculate semantic similarity between each successive item in the letter fluency task by both healthy controls and individuals with schizophrenia. We then subjected these values to a resampling procedure in order to determine whether the observed similarity values were greater than expected by chance. Preliminary results suggest that semantic similarity does indeed play a role in the letter fluency task. Implications for the nature of semantic memory abnormalities in individuals with schizophrenia will also be explored.

3. **The semantic pictionary project**

Michael N. Jones, Brent Kievit-Kylar
Indiana University
jonesmn@indiana.edu, bkivitk@indiana.edu

We describe the NSF Semantic Pictionary Project—a set of online games and tools designed to capture large amounts of structured data about the object characteristics and perceptual properties of word referents. The project hinges on the use of encoding-decoding "games with a purpose" (GWAPs) and a set of creation tools to capture data using online crowdsourcing. We describe the architecture of the basic tools behind the games, the structure of the resulting data, and how this information may be integrated into existing statistical semantic models. We also describe two validations using data collected from one of the tools (2D Geon Pictionary) demonstrating typicality effects in the metrics of raw Geon objects created by subjects, and unique variance in the predictions of word pair metrics over currently used linguistic and property data. The games may be played at www.SemanticPictionary.org.

4. **A computation analysis of semantic structure in bilingual verbal fluency tasks**

Brendan T. Johns, Indiana University, Bloomington
Katherine Young, Christine Sheppard, University of Ottawa, Elisabeth Bruyere
    Research Institute
Michael N. Jones, Indiana University, Bloomington
johns4@indiana.edu

Bilingualism is extremely prevalent, with more than 50% of the world's population being bilingual or multilingual. Recent research has demonstrated differences in cognitive function between bilinguals and monolinguals (see Bialystok, 2009 for a review), but relatively little research has been conducted to explore the exact mechanisms underlying these differences. One active area of research has been exploring the differences in the semantic memory stores of bilinguals. This is often measured with the category fluency task, where subjects are asked to name as many items from a specific category (e.g. animals) as possible. The standard approach to analyzing fluency data is through subjective ratings of clustering within a subject's productions (Troyer, et al., 1997). This reliance on subjective ratings raises concerns, and may indeed be unnecessary due to the advances of computational models of semantic memory. Co-occurrence models of semantics utilize the structure of the linguistic environment to construct representations of word meanings. We will demonstrate how the representations created by these models, combined with a simple decision mechanism (the Luce choice axiom; Luce, 1959), can be used to assess bilingual word processing in a number of different tasks. We analyze fluency in each of the bilingual's languages (French and English), a mixed-language condition, in which participants provided responses in whichever language they preferred, and a forced-switch condition, in which participants switched between French and English (e.g., "cat", "chien" [dog], "frog", etc.). This co-occurrence based

analysis affords a more generalizable fine-grained analysis than what current subjective clustering methods provide.

5. **Content coverage analysis based on DSSPP**
Xudong Huang, Jun Xie, Xiangen Hu, Quan Tang, Douglas A. Starnes
University of Memphis
xuhuang3@memphis.edu, jxie2@memphis.edu, xhu@memphis.edu, quantang@memphis.edu, dstarnes@memphis.edu

The domain-specific semantic processing portal (DSSPP) is a collection of web-services that process texts semantically. The functionality of DSSPP is similar to the popular LSA portal from the University of Colorado (http://lsa.colorado.edu). However, DSSPP is not limited to a single semantic encoding method. DSSPP provides the means to measure the semantic overlap of any two texts from different semantic spaces. Furthermore, the semantic overlap computed from DSSPP is also domain-specific. For example, the semantic overlap between "force" and "acceleration" might be different between the domains of physics and biology. In this paper, we apply DSSPP in content coverage analysis for any written English text within any given domain. There are two steps to implement content coverage analysis using DSSPP: answer-key analysis and semantic overlap analysis. In answer-key analysis, we select a semantic space and domain combination such that the answer keys are the most semantically distinctive. We then use the selected semantic space and domain to compare the semantic overlap between the target essay and answer keys. Using these results, we are able to provide a measure of content coverage in an essay grading system.

6. **A holographic examplar model of implicit learning**
Chrissy M. Chubala, Randall K. Jamieson
University of Manitoba

The way a stimulus is encoded impacts peoples' judgments in the artificial grammar task (Jamieson & Mewhort, 2005; Wright & Whittlesea, 1997). However, few models of learning in the artificial grammar task take encoding into account. To examine the problem closely, we report an artificial grammar experiment. In a training phase, subjects studied letter strings generated according to the rules of a Markov grammar. One group of subjects saw strings presented as successive bigrams, while another saw them presented as successive trigrams. In a test phase, subjects judged the grammatical status of novel letter strings. Subjects in the successive bigram group discriminated test strings' grammatical status, while those in the successive trigram group did not. The difference illustrates an influence of stimulus encoding on judgement of grammaticality. We model subjects' judgments using an exemplar model of memory that borrows its representation scheme from an account of semantic memory (Jones & Mewhort, 2007). In the model, symbols are represented as vectors and strings are constructed by applying noncommutative circular convolution to those vectors. When a probe is presented to memory, it retrieves an aggregate of stored exemplars and judgement of grammaticality is a function of the probe's similarity to

the aggregate. The model tracks performance *only if* stimulus representations are constructed to correspond to the encoded representations that people used. The model integrates a new method for representation in semantic memory with the storage and retrieval operations of an exemplar model of memory.

7. **Modeling corrective saccades**
Pablo Gomez, Christopher Cox, Jason Geller
DePaul University
pgomez1@depaul.edu

Information stored in visual short term memory (VSTM) plays an important role in controlling eye movements. If needed, information about an intended target is utilized immediately following a saccade to make a corrective saccade. In this poster, we present a model of the decision stage in the production of corrective saccades. We present three experiments in which we manipulated the discriminability of the targets and distractors by presenting them in different viewpoints (Experiment 1), by presenting highly similar line drawings (Experiment 2), and by presenting orthographic-neighbors word stimuli. The similarity of the distractor to the target had large effects on the accuracy, but no effect on the latency of the corrective saccade. This pattern of results presents a challenge for current models of perceptual decision making, and suggests that the corrective saccades occur after a ballistic decision making process.

8. **A comparative study of visual-spatial performance with organic molecular representations using eye-tracking**
Mike Whitson, John Williams, Bill Harwood
University of Northern Iowa
whitsonm@uni.edu, john.williams@uni.edu, william.harwood@uni.edu

Visual perception has long been considered the primary tool for understanding and manipulating spatial information in the chemical sciences (Stieff et al., 2005), however molecular representations can include both spatial and verbal information (e.g. letters or numbers). When presented with images, individuals do not always encode these objects as mental images alone, rather visual data can be stored as spatial information and non-spatial (verbal) data can be encoded as propositional statements (Pickering, 2001). Further, in a reciprocal relationship, non-spatial information like numbers, letters, or words can cue individuals to mentally visualize an image of the specific molecular representation (Shepard & Chipman, 1970). Given this reciprocal cooperation between verbal and spatial memory, the current study uses eye-tracking to investigate what features of organic chemistry molecular representations participants examine, what differences exist between novice-intermediate-expert viewers, and cueing methods to improve accurate identification and retention of organic molecules. In addition, differences in identification accuracy between molecular images providing only spatial information (ball & stick, space filling) and images utilizing spatial and non-spatial information (line-drawing) are

investigated to elucidate whether one particular type of representation is retained better in subsequent tests.

9. **An r package for state-trace analysis**
Melissa Prince, Guy Hawkins, Jonathon Love, Andrew Heathcote
The University of Newcastle, Australia
Melissa.Prince@newcastle.edu.au, Guy.Hawkins@newcastle.edu.au,
Jonathon.Love@newcastle.edu.au, Andrew.Heathcote@newcastle.edu.au

State-trace analysis (Bamber, 1979) is a graphical method that can determine whether one or more than one latent variable mediates an apparent dissociation between the effects of two experimental manipulations. State-trace analysis makes only ordinal assumptions, and so is not confounded by range effects that plague alternate methods, especially when performance is measured on a bounded scale (such as accuracy). We describe and illustrate the application of a freely available GUI driven package, *StateTrace,* for the R language. *StateTrace* automates many aspects of a state-trace analysis of accuracy and other binary response data, including implementing Prince, Brown and Heathcote's (2011) Bayesian procedures for quantifying evidence about the outcomes of a state-trace experiment and the creation of graphs.

10. **Multimodal synchrony: tracking body and voice in an affordable behavioral recording setup**
Alexandra E. Paxton, Rick Dale, Cognitive and Information Sciences, University of California, Merced
alon@memphis.edu, rdale@ucmerced.edu

Research on behavioral and vocal interpersonal synchrony during conversation has historically required large investments in equipment and/or labor. The experimental parameters have demanded specialized tools, and data analysis sometimes called for costly training of research assistants, with a hand-coding process that was equally laborious. One goal of the present study is to design an affordable yet efficient alternative that provides access to an array of behavioral signatures to study synchrony during interaction. To illustrate the setup, an experiment is presented that examines interpersonal synchrony during asymmetric interactions, focusing on the differences between argumentative and affiliative interactions. Virtually all studies focus on the affiliative function of synchrony, and these results offer new insight into how synchrony is organized in different forms of interaction. The experiment utilized cheaper alternatives to specialized equipment, including commercially-available recording equipment with a low budget. Similarly, researchers analyzed the data using a series of automated analyses written in MATLAB. By automating the process, the present study minimized the cost of hand-coding and the effort required to analyze it. Reducing the expense of equipment, data collection, and processing may allow those with limited resources to contribute new research to areas that were previously outside their means. Related methodological approaches will be compared,

and this work ends on theoretical discussion of how similar research may shed new light on conversational coordination across multiple modalities.

11. **Impacts of perceptual complexity on target detection performance: A nonlinear mixed model analysis**
Min-Ju Liao, Ching-Fan Sheu, National Chung-Cheng University, Taiwan
mjliao@cc.edu.tw, csheu@mail.ncku.edu.tw

Advanced technology enables us to develop complex interface systems. However, demands from complex interface systems may exceed users' information processing capacity, thereby causing task performance to deteriorate. Xing (2007) proposed a framework for measuring information complexity. In this framework, complexity is composed of three factors, quantity, variety, and relations among information elements, and each factor is evaluated at the three stages of information processing, perception, cognition, and action. The present experiments evaluated the impacts of perceptual complexity on target detection performance within a computer display. Quantity of items and the clutter level of background graphs were manipulated. Participants searched among alternatives within a brief period of time, decided whether or not the target icon was present, and reported the confident level of their responses. Changes of signal detection measures sensitivity (d') and decision criterion (β) under different complexity levels were parameters of interest. Due to considerable individual differences and numerous incidents of zero hit and false alarm rates in the present data, calculations of d' and β based on trials within an individual were no longer suitable. A nonlinear mixed-effect signal detection model (Sheu, Lee, & Shih, 2008) was therefore adopted where subject-specific random variables were introduced to estimate sensitivity and decision criterion for each experimental condition. It was found that participants' sensitivity decreased as the quantity of items and background clutter level increased. The present study demonstrated that a nonlinear mixed-effect approach was more efficient and powerful in estimating d' and β for comparing the experimental effects of interests.

12. **A solution for measuring accurate reaction time to visual and auditory stimuli and its application for assessments in occupational therapy**
Toshio Ohyanagi, Yasuhito Sengoku, Sapporo Medical University
Masako Miyazaki, Lili Liu, University of Alberta
ohyanagi@sapmed.ac.jp, sengoku@sapmed.ac.jp, masako.miyazaki@ualberta.ca, lili.liu@ualberta.ca

We developed a USB device realized with Cypress Programmable System-on-Chip (PSoC) mixed-signal array programmable microcontrollers as a solution for measuring accurate reaction time (SMART). The SMART device was to measure reaction times (RTs) to visual stimuli, and only one touch sensor was used forgetting responses. We have modified the SMART device to measure RTs to auditory stimuli as well as visual stimuli and to add more touch sensors for getting responses. These modifications are necessary for RT tasks used for assessments in Occupational Therapy. Furthermore, we have modified the SMART device in order that the

participants can respond by not only touching their fingers to the sensors but also releasing their fingers from the sensors, since some children and elderly participants had made errors in touching the sensor. In this presentation, we will discuss the modified SMART device and its timing accuracy in measuring RTs. We also present software, termed SMART software, in which we implemented simple RT tasks, a discriminating figures task, discriminating visual and/or auditory target tasks, and other tasks for assessing patients in Occupational Therapy. For some of the tasks, we can set moving images or dots in the background of the display as distractors. The SMART software was developed using REAL Studio software, and then it runs on both Windows and MacOS X environments. We will present our preliminary results of assessing patients using the developed tasks, and we will demonstrate our system.

13. **Behavioural distinction between strategic control and spatial realignment during visuomotor adaptation in a viewing window task**
Jane M. Lawrence, Lee A. Baugh, Jonathan J. Marotta
University of Manitoba
lawrencj@cc.umanitoba.ca, lee@biomed.queensu.ca, marotta@cc.umanitoba.ca

We must frequently adapt our movements in order to successfully perform motor tasks. These visuomotor adaptations can occur with or without our awareness and so, have generally been described by two mechanisms: strategic control and spatial realignment. Strategic control is a conscious modification used when discordance between an intended and actual movement is observed. Spatial realignment is an unconscious recalibration in response to subtle differences between an intended and efferent movement. Traditional methods of investigating visuomotor adaptation often involve simplistic, repetitive motor goals and so may be vulnerable to subject boredom or expectation. Our laboratory has recently developed a novel, engaging computer-based task, the *Viewing Window,* to investigate visuomotor adaptation behaviour. Here, we demonstrate that this paradigm can behaviourally distinguish between strategic control and spatial realignment. Subjects experienced either gradual, progressive rotations or large, sudden rotations in *Viewing Window* movement. The gradual rotation group demonstrated significantly faster mean velocities and scan times, and spent significantly less time off the object compared to the sudden rotation group. These differences demonstrate adaptation to the distortion using spatial realignment. Scan paths revealed greater after-effects in the gradual rotation group reflected by greater time spent scanning areas off of the object. These results demonstrate the ability to investigate both strategic control and spatial realignment. Thus, the *Viewing Window* provides a powerful engaging tool for investigating the neural basis of visuomotor adaptation and impairment following injury and disease.

14. **Finger tracking in video using ridge detection**
Bas H. M. van der Velden, Eindoven University of Technology
Valerie S Morash, University of California, Berkley

Eye tracking is widely used in psychological research. There are many cases in which hand and finger tracking would be very useful. Fiducial markers on the fingers can be robustly tracked with a camera or electrical receiver, but this method requires special equipment, and may limit the subject's freedom of movement. Single-camera machine vision approaches are desirable for their use of non-specialized hardware, and usability in unrestricted environments, but the algorithms for these approaches are not well developed. We have adopted a ridge detection algorithm, previously found useful for medical image analyses, to enable robust finger tracking. Ridge detection is distinct from edge detection. It operates by computing the local second derivative in the direction orthogonal to the local image gradient. We used this method to quantify how long peoples' fingertips are in contact with parts of the stimuli, and how quickly the fingers are moving: analyses analogous to those done using eye tracking. Benefits of our approach are that we do not need fiducial markers or special cameras, and our algorithm is effective in a relatively uncontrolled environment. Our results suggest that ridge detection, used in combination with previous approaches such as edge detection and hue segmentation, could provide psychological researchers with an effective tool for hand and finger tracking.

15. **Demonstrating cognitive experiments using PsyToolkit and Java**
Gijsbert Stoet
Institute of Psychological Sciences, University of Leeds
g.soet@leeds.ac.uk

PsyToolkit is a software package for programming and running psychological experiments typical in cognitive psychology. Originally, the software only ran on the Linux operating system, but recently a Java version has been added. Java has various advantages. It runs on all common platforms, it is freely available, can be installed easily, and can be started from a web browser using the Java web start technology. Using PsyToolkit for Java, a set of lessons and demonstrations about a number of common experiments in cognitive psychology has been created: http://psytoolkit.leeds.ac.uk/lessons (although the Java version works well enough to demonstrate a variety of common cognitive phenomena, it is not yet recommended for time-critical experiments). The lessons are written in simple language to make them easily accessible, not only for university students, but also for high-school students and the interested public. Further, the lessons and demonstrations come with detailed instructions on how to analyze resulting data using the free statistical software R and with instructions on how to create stimuli using the free software Inkscape. The combined use of freely available software for demonstrating experiments lowers the barrier for experiencing and experimenting with cognitive experiments, and will potentially lead to a larger interest in cognitive psychology.

16. **DMTpsych: Postgraduate training for research data management in the psychological sciences**
Richard R. Plant, Philip Quinlan, The University of York, UK
Will Reader, Sheffield Hallam University, UK
Andrew Thompson, The University of Sheffield, UK
r.plant@psych.york.ac.uk, p.quinlan@psych.york.ac.uk, w.reader@shu.ac.uk,
A.R.Thompson@sheffield.ac.uk

Today's experimental psychology research produces large amounts of raw data, much of which is generated by or stored on state-of-the-art computers or other machinery. However, the pace of technological development means that it can be very difficult to access raw data from an experimental psychology study that took place only a decade or so ago - even though it is still possible to access the raw data from sun observations made by Copernicus more than 500 years ago. In the US, the National Science Foundation is addressing issues of data retention and reuse with guidance on how data should be made available for up to 100 years. In the UK, research councils including the ESRC, BBSRC and Wellcome are making data-handling a core requirement of funding, and universities are increasingly implementing their own data handling requirements. However, it has also been recognised that different disciplines often require discipline-specific guidance on how best to store their data. The JISC funded DMTpsych project aims to provide such guidance for the psychological sciences. In so doing, the project aims to enable postgraduates in the psychological sciences to develop a sustained awareness of the importance of data organisation, validation, sharing and curation, and to make such an awareness a 'core academic competency'. This paper will discuss theses issues and the cultural change they may require, and will also outline how the postgraduates of today and the researchers of tomorrow may benefit from DMTpsych.

17. **Space fortress 5**
Ryan M. Hope, Marc Destefano, Michael Schoelles, Wayne Gray
Rensselaer Polytechnic Institute
hoper2@rpi.edu, destem@rpi.edu, schoem@rpi.edu, grayw@rpi.edu

This article presents Space Fortress 5, the latest manifestation of the vector-based, friction-less, space themed video game initially developed by Mane & Donchin (1989) as part of the DARPA funded "Learning Strategies Project". The subtasks of Space Fortress are one-o_s of many popular cognitive science experimental paradigms (ex: N-Back, AX-CPT and Sternberg memory) integrated into a single complex task environment. This makes Space Fortress an excellent research tool for studying a number of topics of interest to cognitive scientists such as cognitive control, skill acquisition and strategy discovery. However, the infrequent and minimal logging of previous versions of Space Fortress prevents it from being of use for anything but the most basic research. Space Fortress 5 is open source and written in Python/PyGame, allowing it to run on Mac OS, Windows and Linux. Every aspect of Space Fortress 5 is customizable via a new GUI con_guration editor. Other major features of Space Fortress 5 include millisecond level logging and a bi-directional

plugin framework that allows the game to interact with hardware (like eye trackers, EEG and fMRI) and cognitive architectures (such as ACT-R).

18. **Using web games to elicit associative and feature-based conceptual representations**

Gabriel Recchia, Brent Kievit-Kylar, Michael Jones, Indiana University
Ken McRae, University of Western Ontario
grecchia@indiana.edu, bkievtik@umail.iu.edu, jonesmn@indian.edu, kenm@uwo.ca

Although researchers have long considered associative and semantic relatedness to be distinct constructs, this dichotomy has been challenged with evidence suggesting that associative relatedness may be a byproduct of semantic relations (McRae, Khalkhali, & Hare, 2011). If this is the case, one might expect predictors of associative and semantic relatedness to tap into some of the same psychological characteristics. For example, cue set size (QSS), the number of associates that at least two participants produced in response to a given cue word in the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998), and number of features (NF), the number of features that at least 5 of 30 participants produced for a word's referent in McRae, Cree, Seidenberg, & McNorgan's (2005) semantic feature production norms (Pexman, Lupker, & Hino, 2002), may be measuring the same underlying property of semantic relations. Even though QSS and NF are relatively isomorphic measures, direct comparisons between predictors of semantic and associative relatedness are rarely straightforward due to differences in the way most measures are defined in the two sets of norms. We use a web-based game to collect a database of features and a database of associates for over 200 words that correlates well with the McRae and Nelson norms, respectively. We release a rich dataset with information lacking from these norms (order of responses, number of responses per participant, etc.) and which allows us to directly compare predictors of semantic and associative relatedness in a more controlled fashion.

19. **Considerations for online homework systems**

Christopher Koch
George Fox University
ckoch@georgefox.edu

Online homework systems allow instructors to organize assignments and support materials to facilitate learning. These systems can send reminders about upcoming assignments, include deadlines for completing assignments, and provide additional problems for students who desire extra practice with the material. Such a system was adopted for a college-level statistics course. From an instructor's perspective, the system had a number of strengths and weaknesses. For instance, assigning problems is a fairly easy process as long as the adopted text also has a set of problems ready for the system. If such a problem set is not available, the instructor must create a problem set which can be time consuming. Furthermore, in order to benefit from the automatic scoring, only problems that require a fixed response can be assigned. Problems requiring a written response need to be scored manually. Scoring

essentially occurs on an all or none basis. Partial credit is difficult to assign. Depending on the grading rubric, this limitation can lower assignment grades. Students, on the other hand, were generally frustrated by the system. The two most common complaints dealt with time and access. First, students had to enter responses online that they had already written down. They felt that this form of double entry unnecessarily increased the amount of time required to complete the assignments. Second, the system was a stand-alone system. Therefore, students needed to visit an additional site that they were not accustomed to using. Additional considerations for online homework systems are discussed.

20. **Video games and visual search: The roles of competition and experience**
Patrick Conley, Jonathon M. Wutke, Denise Calhoun, Ian K. Evans
University of Wisconsin-Stevens Point
pconley@uwsp.edu, Jonathon.M.Wutke@uwsp.edu, Denise.Calhoun@uwsp.edu, Ian.K.Evans@uwsp.edu

This study extends previous work on the effect of an individual's video gaming history on visual search task performance. In previous work (Conley et al., 2010), the authors demonstrated that various aspects of an individual's video gaming experience (age of exposure to video games, amount of video gaming, and having other active video game players in the household) successfully predicted visual search task performance, with more and earlier video gaming leading to superior performance in the search task. This work extends the previous work by greatly expanding the video gaming questionnaire that was administered to the experimental participants. In addition to the questions included in the previous questionnaire, the roles of other video gaming family members, genre of the video games, and levels of competitiveness were also explored. The results of the current research revealed a similar pattern to the previous study, but also demonstrated that the type of video games played had an effect on performance, with more action-oriented games leading to better performance in a visual search task. In addition to the basic experimental result, the results of the video gaming questionnaire will be discussed to determine what they reveal about the role of the video games in the past and current lives of college students.

21. **Balance beam aid for instruction in medical diagnosis**
Robert M. Hamm, William H. Beasley
University of Oklahoma Health Sciences Center

Understanding how clinical findings discriminate among possible causes of a patient's presentation is essential in clinical judgment, but students start with little intuition about these issues. The balance beam offers a physical analogue for assessing how patient symptoms impact the relative strength of competing disease hypotheses. We describe a Balance Beam Aid for Instruction in Diagnosis (BBAID) for training medical students to diagnose a common disease presentation. The BBAID lets the student place competing diseases on each end a balance beam; weights representing evidence are placed at appropriate locations on the beam. This familiar

physical analogue can accurately represent the combination of pieces of evidence with different discrimination abilities. The impact of patient findings, expressing Bayes' theorem with log odds and log likelihood ratios, maps directly to the forces affecting a balance beam. With the BBAID, the user specifies whether each finding is present, absent, or unknown. For known findings, tokens are placed on the beam at locations whose distance from the fulcrum is proportional to the finding's log(LR). Positive findings get weights (downward force); negative findings get balloons (upward force). The summary impact of all evidence is displayed. A separate matrix shows the current probabilities of each disease and their pairwise comparisons. Clicking a cell in this matrix causes the corresponding pair's data to be displayed on the balance beam and LR's for that pair to be displayed next to the symptom names. The BBAID is a tool for research, education, or clinical practice.

22. **Web-based problem-solving for teens with traumatic brain injury: Who benefits?**

Shari L. Wade, Cincinnati Children's Hospital Medical Center
H. Gerry Taylor, Rainbow Babies & Children's Hospital and Case Western Reserve University
Terry Stancin, MetroHealth Medical Center and Case Western Reserve University
Michael Kirkwood, The Children's Hospital of Colorado
Tanya M. Brown, The Mayo Clinic
Shari.wade@cchmc.org, hgt2@cwru.edu, tstancin@metrohealth.org, Michael.Kirkwood@childrenscolorado.org, Brown.tanya@mayo.edu

The incidence of traumatic brain injury (TBI) peaks in adolescence. Its effects are varied and include deficits in executive function skills, such as planning and problem solving, changes in behavior, and social and communication difficulties. Given that adolescents are native technology users, they may be particularly able to benefit from treatments delivered via the internet. This study examines which teens with TBI experience the greatest reductions in executive dysfunction and behavior problems following Counselor Assisted Problem Solving (CAPS), a web-based problem-solving therapy. Teen age, injury severity, verbal IQ, and socioeconomic status were examined as potential moderators of treatment response. One hundred thirty two adolescents, ages 12-17 with moderate to severe TBI, and their families were randomly assigned to receive CAPS or internet-based resources regarding TBI (Internet Resource Comparison; IRC) and executive function behaviors and behavioral problems were assessed prior to treatment and again 6-months later. To bridge the digital divide, all participants received a computer to keep and high speed internet access for the duration of the interventions. Adolescents in both groups were an average of 14 ½ years old and 3.6 months post injury. Multiple regression analyses revealed overall improvements in the CAPS versus IRC group for some domains of behavior. Additionally, adolescents who were older and those with lower verbal intelligence were more likely to benefit from CAPS than IRC. These findings suggest that adolescents with TBI can benefit from a web-based problem solving approach and raise the possibility of targeting those who are most likely to benefit.

23. **Computer versus faculty generated assessment of surgical simulation performance: What are the differences?**

Angela Brunstein, Joerg Brunstein, Carnegie Mellon University in Qatar
Sandeep Reddy Pamulapati, Northwestern University in Qatar
Bakr Nour, Weill Cornell Medical College in Qatar
Davit Sargsyan, Hamad Medical Corporation
angelab@cmu.edu, joerg@brunstein.net,
sandeeppamulapati2014@u.northwestern.edu, bmn2001@qatar-med.cornell.edu,
sargdav@yahoo.com

Recently, surgical simulation training has become mandatory for residency in the US. Surgical simulators provide for assessing performance during training objective evaluation and standardized cases. In this study, we compared automatically generated assessment from the simulator with faculty's assessment for randomized and anonymized screen videos on laparoscopic cholecystectomy for pre- and post test and for a transfer case. 14 medical students after surgical internship practiced the procedure for 5 sessions of 30 min each using a VR simulator (CAE Healthcare). The simulator's protocols included a pass / fail grade and checkmarks for proficiency, dexterity, and complications. Faculty's assessment resulted in a total score consisting of instrument handling, dissection, and gallbladder removal, minus penalties for damage. The inter-rater reliability for the 2 faculty members was comparably high (α= .91). However, the concordance between the simulator's and faculty's pass/fail grades (pre: 94%, post: 61%, transfer: 67%) and for total scores was comparably low (pre: -.18, post: .28, transfer: -.07). For the protocols, checkmarks for time and proximity did not discriminate between participants and trials. In contrast, proficiency scores and complications varied between participants and improved with practice. Faculty's assessment was more fine-grained and prioritized some parts of the procedure before others. The protocols seem to encourage more speed improvements, while faculty seems to encourage accuracy improvements. Therefore, current surgical simulation systems are great tools for practicing surgical procedures in a safe and encouraging environment, but are not efficient yet for providing trainees with meaningful feedback and evaluation.

24. **Toward automation of verbal protocol data analysis**

Shahram Ghiasinejad
University of Central Florida
ghiasinejad@ucf.edu

Free response data analysis usually involves identifying the propositions that are explicitly mentioned in a written or oral report for subsequent analysis. Typically, two experienced coders independently semantically annotate the protocol data and then agreement measures are used to determine the reliability of the coding. However, human coding is subjective, not perfectly replicable, and is often poorly documented. These characteristics make it difficult to communicate the coding procedures in an entirely unambiguous manner to other researchers. We have been exploring an automatic coding methodology as an alternative to manual coding of free response

data which we call AUTOCODER. The AUTOCODER system works by interacting with an experienced human coder who semantically annotates key words with "word concepts' and sequence of words with "propositions". After training, AUTOCODER automatically segments and semantically annotates novel protocol data exhibiting a good proposition agreement rate of 91% and a good kappa agreement score of 65% with respect to an experienced human coder. Finally, strengths, limitations, and distinctive characteristics of the AUTOCODER methodology are discussed.

25. **Rethinking the automatic assessment of reading processes: Is it just a matter of doing more?**
Chutuma Boonthum, Old Dominion University
Joseph Magliano, Northern Illinois University
McCarthy…
cboont@cs.odu.edu, jmagliano@niu.edu,

It is well understood that comprehension emerges as the result of inference and strategic processes that support the construction of a coherent text representation. We have been actively involved in developing computer-based tools that provide assessments of the extent that readers engage in these processes. Given the serious challenges involved in automatic natural language processing, we have typically adopts relatively simplistic algorithms that either provide a uni-dimensional assessment of strategic processing or assessments of the extent that readers engage in a small set of processes (e.g., paraphrasing, bridging, elaboration). These approaches ignore the fact that successful comprehenders dynamically use a combination of these strategies to support comprehension. Using the Reading Strategy Assessment Tool (RSAT), we explore the extent readers engage in multiple strategic processing during reading is indicative of comprehension skill. Our data show that the extent that readers engage in single strategies when thinking aloud (e.g., only paraphrasing) is negatively correlated with comprehension skill, whereas the extent that they engage in multiple strategies is positively correlated with comprehension skill, and in particular strategy combinations that involve bridging processes. These results have important implications for RSAT and other computer-based assessments of verbal protocols.

26. **Automatic evaluations of advertising messages**
Christelle Chrea, R&T Chanel PB, Neuilly sur Seine, France
Ching-Fan Sheu, Shu-Hiu Wang, Institute of Education, National Cheng Kung University, Taiwan
Séverine Navarro, Danielle Mougin, R&T Chanel PB, Neuilly sur Seine, France
Christelle.chrea@chanel-corp.com, csheu@mail.ncku.edu.tw,
swang@mail.ncku.edu.tw, severine.navarro@chanel-corp.com,
danielle.mougin@chanel-corp.com

Research has shown that physical movements are related to automatic evaluations, which can be linked to approach or avoidance tendencies. This implicit behavior-related motor action, called *approach –avoidance task* (AAT), has been used broadly

to study phobia or social anxiety but it has not been applied in consumer research. In this work, we examined attitudes towards advertising messages by employing the Approach-Avoidance Task. A sample of 80 female participants from two age groups (30 to 45 and 50 to 65 years) responded to a series of brief advertising messages in reference to anti-aging cosmetic efficacy by pulling a joystick towards themselves (positive approach movement) or by pushing it away from themselves (negative avoidance movement). Following the task, participants completed the five appearance-related items of the *anxiety about aging scale* (Lasher & Faulkender, 1993). No age effect on stimuli evaluations was observed, however, a regression analysis of the joystick movements revealed a significant relationship between anxiety about aging and the valence of the messages (positive or negative). To product developers and marketers, the findings provide useful insight for describing products so as to better meet consumer needs and expectations.

[Vendor Page]