SCiP 2018

Officers of SCiP

President: Chris Westbury, *University of Alberta*
Executive Director: Geoff Hollis, *University of Alberta*
President Elect: Danielle McNamara, *Arizona State University*
Past President: Xiaowei Zhao, *Emmanuel College*

Steering Committee

Laura Allen, *Mississippi State University*
Blair Armstrong, *University of Toronto*
Pietro Cipresso, *Catholic University of Milan*
Carol Connor, *Arizona State University*
Brendan Johns, *University at Buffalo*
Danielle McNamara, *Arizona State University*
Alexandra Paxton, *University of Connecticut*
Ulf-Dietrich Reips, *University of Konstanz*
Stefan Stieger, *University of Konstanz*
Joseph Young, *General Dynamics Information Technology*

[Cover image: Computer-altered open source photograph of the great New Orleans musician, Louis Armstrong.]
Welcome from the President

Welcome to the 48th annual meeting of the Society for Computers in Psychology! It has been my pleasure to oversee the formation of this year’s program, with a great deal of help from SCiP’s new executive director, Geoff Hollis, as well as from our former executive director Rick Dale.

For the President’s Symposia I chose the theme One law to rule them all: Statistical learning as psychological explanation. I am very pleased to have as a keynote speaker Harald Baayen from the University of Tübingen, whose computational work in modeling language in recent years serves as a systematic lesson in how to be skeptical about received wisdom on the fundamental units of language processing, a lesson which has implications for how we might think about other psychological functions. His talk will address this, as he is speaking on: Throwing off the shackles of the morpheme with simple linear mappings. The president’s symposium includes three other people whose work I admire very much. Jon Willits (University of Illinois) will speak on: The future of deep learning models of language and semantics. Marc Brysbaert (University of Ghent) will speak on: Big data in language research: New findings, new challenges. Paul Thagard (University of Waterloo) will close the symposium, speaking on: How does human AI measure up to human intelligence? Thanks to all our speakers for making the effort to come and speak to us at SCiP, as we ask much of them and offer little.

Thanks to Ulf-Dietrich Reips for organizing a symposium this year, and to the vendors for their support. Please stop by their tables while at the conference.

In recognition of the strange times in which we are meeting, I have used my vast presidential powers to rename all the SCiP 2018 coffee breaks ‘Covfefe breaks’.

Thank you all for joining us in New Orleans. Please enjoy the talks and posters, as well as the great music and great food outside in this wonderful city.

Chris Westbury
University of Alberta
<table>
<thead>
<tr>
<th>Time</th>
<th>Session 1: Experimental tools</th>
<th>Session 2: Language I</th>
<th>Session 3: Experimental Work I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Room: Strand 7</td>
<td>Room: Strand 8</td>
<td>Room: Strand 11A</td>
</tr>
<tr>
<td>8:15</td>
<td><strong>Chair: Blair Armstrong</strong></td>
<td><strong>Chair: Matt Crump</strong></td>
<td><strong>Chair: Kay Livesay</strong></td>
</tr>
<tr>
<td>8:15</td>
<td>Mousetrap: Open-source tools for advanced analyses of hand- and mouse-tracking data</td>
<td>Suggesting Keywords for Automated Discourse Analysis Using LSA Nearest Neighbors</td>
<td>Testing the limits of transferability</td>
</tr>
<tr>
<td>8:30</td>
<td>Creating tools for semantic space visualization</td>
<td>Instances and Information: An instance theory account of sensitivity to uncertainty in natural text production</td>
<td>Using Machine Tools to Analyze Students’ Ethics Essays</td>
</tr>
<tr>
<td></td>
<td>Erin M. Buchanan, Amber Gillenwaters, William E. Padfield, Addie Wikowsky, Arielle Cunningham</td>
<td>Matthew J. C. Crump, Nicholaus P. Brosowsky</td>
<td>Roman Taraban, Lakshmojee Koduru, Mark LaCour, John R. Schumacher, Micah Iserman, Taleen Nalabandian, Joseph A. Martis, Rogelio Carrillo</td>
</tr>
<tr>
<td>8:45</td>
<td>Who said browser-based experiments can't have proper timing? Implementing accurate presentation and response timing in the browser</td>
<td>A Method for Automatically Evaluating Texts for Gist Inferences</td>
<td>Validating Instagram Likes as a Measure of Aesthetic Liking</td>
</tr>
<tr>
<td></td>
<td>Felix Henninger, Yury Shevchenko, Ulf K. Mertens, Pascal Kieslich, Benjamin E. Hilbig</td>
<td>Christopher R. Wolfe, Mitchell Dandignac, Valerie F. Reyna</td>
<td>Katja Thömmes, Ronald Hübner</td>
</tr>
<tr>
<td>9:00</td>
<td>A free and open-source toolkit of three-dimensional models and software to study face perception</td>
<td>Role of Picture-Text Correspondence in Perceived Accuracy of Online Fake News</td>
<td>That Emoji Means What? Exploring the Semantics of Emoji</td>
</tr>
<tr>
<td></td>
<td>Fabian A. Soto, Jason Hays, Claudia Wong</td>
<td>Aiping Xiong, Robert W. Proctor</td>
<td>Kay L. Livesay, Megan Kozak Williams</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Time</th>
<th>Covfefe Break I</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:15</td>
<td></td>
</tr>
</tbody>
</table>
| 9:30 | Session 4: Applied Computation  
Room: Strand 7  
Chair: Randy Jamieson | Session 5: Language II  
Room: Strand 8  
Chair: Brendan Johns | Session 6: Symposium: New methods and tools in Internet-based research  
Room: Strand 11A  
Chair: Ulf-Dietrich Reips |
|---|---|---|---|
| 9:30 | Digital Traces of Identity in Online Discourse: Detecting Socio-Cognitive Roles with Group Communication Analysis  
Nia Dowell, Christopher Brooks, Oleksandra Poquet | Writing for Coh-Metrix: A Systematic Approach to Revising Texts to Foster Gist Inferences  
Mitchell E. Danzignac, Christopher R. Wolfe | Acquiescence in Internet-based research: An artifact of socially desirable responding?  
Vaka Vésteinsdóttir, Ragnhildur L. Asgeirsdottir, Fanney Thorsdottir, Ulf-Dietrich Reips |
| 9:45 | The lifecycle of open-source software: Mining GitHub to understand community dynamics  
Alexandra Paxton, Nelle Varoquaux, R. Stuart Geiger, Chris Holdgraf | The effects of time and place on lexical behavior: A distributional analysis  
Brendan T. Johns, Randall K. Jamieson | Guidelines for improving the reliability of Google Ngram studies: Evidence from religious terms  
Nadia Younes, Ulf-Dietrich Reips |
| 10:00 | Simulating Dynamic Task Redistribution within Human-Machine Teams Using an Autonomous Manager  
Mary E. Frame, Jennifer S. Lopez, Alan S. Boydstun | Deep Learning of the basic structures of Chinese characters  
Xiaowei Zhao | Improving data quality in Internet-based studies: CRT item generator  
Deborah Canales, Pablo Garaizar, Ulf-Dietrich Reips |
| 10:15 | RotCorr and MDSat: New ways to optimize rotation of multidimensional scaling solutions  
K. Jake Patten, Michael K. McBeath | Using large text corpora to compute word emotionality scores: Construction and validation of a corpus-derived Chinese word emotionality database  
Yen-Cheng Chen, Meng-Ning Tsai, Yao-Ting Sung, Yu-Lin Chang, Shu-Yen Lin, Shu-Ling Cho | Methods for Studying Sensitive Topics Online  
Ulf-Dietrich Reips, Deborah Canales |
| 10:30 | Using semantic memory theory to build an applied cognitive search engine  
Randall K. Jamieson, Harinder Aujla, Matthew J. C. Crump, Matthew T. Cook | Structure of scale-free random networks  
Peizhu Qian, Zhiqiang Cai, Xiangen Hu | Smartphone sensor data in experience sampling: The gyroscope and its heterogeneities  
Tim Kuhlmann, Pablo Garaizar, Ulf-Dietrich Reips |

Discussant: Chris Wolfe
<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
<th>Presenters</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:45</td>
<td><strong>Knowledge Structures for Medication Side Effects: Human vs. Computational Methods</strong></td>
<td>Ruth S. Day, Jiehan Zhen</td>
</tr>
<tr>
<td>11:00</td>
<td><strong>Covfefe Break II</strong></td>
<td></td>
</tr>
<tr>
<td>11:15</td>
<td><strong>Session 7: Experimental Work II</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Room: Strand 11A</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Chair: Caitlin Mills</strong></td>
<td></td>
</tr>
<tr>
<td>11:15</td>
<td><em>An Improved Assessment Model for AutoTutor Answers</em></td>
<td>Keli Cheng, Zhiqiang Cai, Art Graesser</td>
</tr>
<tr>
<td>11:30</td>
<td><em>A Real-time Mind Wandering Intervention during Reading</em></td>
<td>Caitlin Mills, Julie Gregg, Robert Bixler, Sidney D'Mello</td>
</tr>
<tr>
<td>11:45</td>
<td><em>Computer- and Paper-Based Normative Ratings of Metaphors: They are Similar in Some Ways and Different in Other Ways</em></td>
<td>Cesar E. Riano Rincon, Gary E. Raney</td>
</tr>
<tr>
<td>12:00</td>
<td><strong>Lunch &amp; Poster setup [Room: Elite Ballroom on the first floor.]</strong></td>
<td></td>
</tr>
<tr>
<td>1:00</td>
<td><strong>Poster session [Room: Elite Ballroom on the first floor.]</strong></td>
<td></td>
</tr>
<tr>
<td>2:15</td>
<td><strong>Presidential symposium [Room: Strand 11A]</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Jon Willits: The future of deep learning models of language and semantics</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Marc Brysbaert: Big data in language research: New findings, new challenges</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Paul Thagard: How does current AI measure up to human intelligence?</em></td>
<td></td>
</tr>
<tr>
<td>3:45</td>
<td><strong>Covfefe break III</strong></td>
<td></td>
</tr>
<tr>
<td>4:00</td>
<td><strong>Keynote talk [Room: Strand 11A]</strong></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td>5:15</td>
<td>Business meeting</td>
<td></td>
</tr>
</tbody>
</table>

Harald Baayen: *Throwing off the shackles of the morpheme with simple linear mappings*
President Symposium
Strand 11A

One law to rule them all:
Statistical learning as psychological explanation

In recent decades, psychologists have seen computational developments (especially as they relate to our ability to collect and understand big data sets) change the very way we think of psychological explanation. Many ontological assumptions that have underlain psychological thinking for many decades are now being cast into doubt by our access to the large datasets that allow for new ways of modeling, thinking about, and explaining psychological phenomena. In this symposium we address some of these changes and the challenges and opportunities they present for psychology.

Jon Willits
*University of Illinois*
**The future of deep learning models of language and semantics**

Marc Brysbaert
*University of Ghent*
**Big data in language research: New findings, new challenges**

Paul Thagard
*University of Waterloo*
**How does current AI measure up to human intelligence?**
Keynote Address
Strand 11A

Harald Baayen
University of Tübingen

Throwing off the shackles of the morpheme with simple linear mappings
Chair: Blair Armstrong, *University of Toronto*

8:15

**Mousetrap: Open-source tools for advanced analyses of hand- and mouse-tracking data**

Pascal J. Kieslich¹, Dirk U. Wulff², Felix Henninger³, Jonas M. B. Haslbeck⁴, Michael Schulte-Mecklenbeck⁵

¹University of Mannheim ²University of Basel ³University of Koblenz-Landau ⁴University of Amsterdam ⁵University of Bern

The recording and analysis of hand movements has become a popular tool in the cognitive sciences to investigate the temporal development of preferences and the competition between response alternatives. In a typical hand- or mouse-tracking study, participants decide between different options that are presented as buttons on a screen. During this task, their hand movements are continuously recorded by registering the position of the hand directly via motion tracking or indirectly by tracking computer mouse movements. The resulting movement data can be analyzed in a multitude of ways to shed light on the underlying cognitive processes and test psychological theories. In this contribution, we present a number of open-source tools that implement recent advances in the analysis of mouse- and hand-tracking data. These include tools for the spatial clustering of trajectories, the mapping of trajectories onto prototypes, and assessing the temporal order with which different areas of interest were visited in a trial. Besides, we offer tools for visualizing movement trajectories via animations, heatmaps, and riverbed plots. These analyses and visualizations allow for answering different research questions, such as which option was initially preferred in a trial, determining if and how often participants changed their mind, and uncovering if movement trajectories are homogenously distributed across trials or whether different types of movements are present in the data. All of these methods are implemented in the mousetrap R package, which is freely available from http://pascalkieslich.github.io/mousetrap/.

8:30

**Creating tools for semantic space visualization**

Erin M. Buchanan, Amber Gillenwaters, William E. Padfield, Addie Wikowsky, Arielle Cunningham

*Missouri State University*

Semantic spaces are used as a representation of language, capturing the meaning between linguistic units. These spaces are often built in large corpora requiring advanced equipment, specialized computational skills, and considerable effort. This presentation will introduce and demonstrate the use of an accessible Shiny graphical interface allowing users to create semantic space models easily. Shiny is an R package in which one can program interactive web applications in R for others to interact with data or analyses. The advantage to Shiny
applications is that naïve users can explore data without understanding the programming, and open sharing of code with the application can aid in learning the programming for one’s own use in their research. Within the application, users will be able to load popular semantic spaces or their own corpus for semantic space creation utilizing their preferred modeling technique, including LSA, TOPICS, HAL, and more. A variety of user-friendly graphical tools, such as n-nearest neighbors or one-hop networks, will further aid data visualization of the semantic network. Additionally, the application provides the calculation of cosine or simple co-occurrence, among other popular-relatedness values. This tool is intended for researchers who may not be programming-savvy, or as a teaching extension for psycholinguistics courses.

8:45

Who said browser-based experiments can't have proper timing? Implementing accurate presentation and response timing in the browser

Felix Henninger¹, Yury Shevchenko², Ulf K. Mertens³, Pascal Kieslich², Benjamin E. Hilbig¹

¹University of Koblenz-Landau, ²University of Mannheim, ³University of Heidelberg

Browser-based studies have been thought (and shown) to limit fidelity of presentation timing and response time measurement, introducing both systematic lags and adding measurement noise. Based on our experience implementing lab.js, a toolkit for online experimentation that aims to provide high-performance stimulus presentation and high-resolution response timing, we demonstrate that these inaccuracies can be minimized, resulting in a vastly improved timing accuracy and precision across browsers. We describe and evaluate strategies for constructing high-performance browser-based experiments, and show that, with careful programming, the limits of browser-based timing are substantially less imposing than previously thought.

9:00

A free and open-source toolkit of three-dimensional models and software to study face perception

Fabian A. Soto, Jason Hays, Claudia Wong

Florida International University

A problem in the study of face perception is that results can be confounded by poor stimulus control. Ideally, experiments should precisely manipulate facial features under study and tightly control irrelevant features. Software for 3D face modeling provides such control, but there is a lack of free and open source alternatives specifically created for face perception research. Here, we provide such tools by expanding the open-source software MakeHuman. We present a database of 27 identity models and 6 expression pose models (sadness, anger, happiness, disgust, fear, and surprise), together with software to manipulate the models in ways that are common in the face perception literature, allowing researchers to: (1) create a sequence of renders from interpolations between two or more 3D models (differing in identity, expression, and/or pose), resulting in a “morphing” sequence; (2) create renders by
extrapolation in a direction of face space, obtaining 3D “anti-faces” and similar stimuli; (3) obtain videos of dynamic faces from rendered images; (4) obtain average face models; (5) standardize a set of models so that they differ only in facial shape features, and (6) communicate with experiment software (e.g., PsychoPy) to render faces dynamically online. These tools vastly improve both the speed at which face stimuli can be produced and the level of control that researchers have over face stimuli. We show examples of the multiple ways in which these tools can be used in face perception research and describe human ratings of stimuli produced with the toolkit.

<table>
<thead>
<tr>
<th>8:15-9:15 AM</th>
<th>Session 2</th>
<th>Language I</th>
</tr>
</thead>
</table>

Chair: Matt Crump, *Brooklyn College of CUNY*

8:15

**Suggesting keywords for automated discourse analysis using LSA nearest neighbors**

Zhiqiang Cai¹, Zachari Swiecki², Brendan Eagan², Xiangen Hu¹,³, Art C. Graesser¹, David W. Shaffer²,⁴

¹*The University of Memphis*, ²*University of Wisconsin-Madison*, ³*China Central Normal University*, ⁴*Aalborg University-Copenhagen*

Educators and researchers who analyze discourse from learning contexts increasingly rely on automated analyses of text data. One use case is the development of classifiers that label discourse events, such as turns of talk, based on their semantic content. Such classifiers can be implemented with keyword lists; however, these users may find it difficult to generate lists that thoroughly cover the semantic content they want to identify—that is, they can generate a short list, but may leave out many important words. A tool that suggests words that are highly semantically related to a short word list would thus benefit these users. This paper investigates three LSA based Nearest Neighbor algorithms, including vector sum, subspace and maximum similarity. The vector sum algorithm sums the LSA vectors of given words and finds nearest neighbors using the LSA cosine of word vectors to this combined vector. The subspace algorithm uses the given word vectors to span a subspace and find nearest neighbors according to the vector projection to the subspace. The maximum similarity algorithm uses the maximum similarity of a word to each of the given words to find nearest neighbors. These algorithms were evaluated using word lists from prior studies of student discourse. The results show that (1) the maximum similarity algorithm provides the best suggestions and takes the shortest time when the nearest neighbors of each word are cached; (2) that a domain specific LSA space out performs a general LSA space; and (3) optimal rare word filtering improves performance.
Instances and information:
An instance theory account of sensitivity to uncertainty in natural text production

Matthew J. C. Crump¹, Nicholaus P. Brosowsky²

¹Brooklyn College of CUNY, ²Graduate Center of CUNY

How does prior experience shape skilled performance in structured environments? We use skilled typing of natural text to evaluate correspondence between performance (keystroke timing) and structure in the environment (letter uncertainty). We had ~350 typists copy-type English text. We reproduced Ostry's (1983) analysis of inter-keystroke interval as a function of letter position and word length, which showed prominent first-letter and mid-word slowing effects. We propose a novel account that letter position and word length effects on keystroke dynamics reflect informational uncertainty about letters in those locations, rather than resource limited planning/buffering processes. We computed positional uncertainty for letters in all positions of words from length one to nine using Google's n-gram database. We show that variance in inter-keystroke interval by letter position and word length tracks natural variation in letter uncertainty. Finally, we provide a model showing how a general learning and memory process could acquire sensitivity to patterns of letter uncertainty in natural English. In doing so, we draw an equivalence between Logan's (1988) instance theory of automatization and Shannon's measure of entropy (H) from information theory: Instance theory's predictions for automatization as a function of experience follow exactly the uncertainty in the choice set being automatized. As a result, instance theory stands as a general process model explaining how context-specific experiences in a structured environment tune skilled performance.

A method for automatically evaluating texts for gist inferences

Christopher R. Wolfe, Mitchell Dandignac, Valerie F. Reyna

¹Miami University, ²Cornell University

We developed a method to automatically assess texts for features that help readers produce gist inferences. Following Fuzzy-Trace Theory, the method was developed using a memory study, in which 64 participants recalled events under gist or verbatim instructions. Using Coh-Metrix, we analyzed their written responses to create Gist Inference Scores (GIS), an average of seven variables converted to z-scores, which is designed to assess the potential for readers to form gist inferences from observable text characteristics. Coh-Metrix measures of referential cohesion and deep cohesion increase GIS because they facilitate forming coherent mental representations. Word concreteness, hypernymy for nouns and verbs (specificity), and imageability decrease GIS because they promote verbatim representations. The difference between abstract verb overlap among sentences using Latent Semantic Analysis minus concrete verb overlap within the same synonym set also enhances coherent gist inferences rather than verbatim memory for specific words. In the memory study, gist-condition responses scored over one standard deviation (significantly) higher on GIS than verbatim-condition responses. Predictions based on the GIS formula were confirmed in two text analysis studies of 50 scientific journal articles and 50 news articles and editorials. An additional
analysis of 244 National Cancer Institute texts about cancer treatments indicates that such texts are unlikely to support meaningful gist inferences, making them less comprehensible than some might assume based on traditional Flesch-Kincaid reading grade level. These findings support a relationship between GIS and readers’ comprehension. We suggest a series of future studies and discuss practical and theoretical implications of Gist Inference Scores.

9:00

Role of Picture-Text Correspondence in Perceived Accuracy of Online Fake News

Aiping Xiong\textsuperscript{1}, Robert W. Proctor\textsuperscript{2}

\textsuperscript{1}Penn State University, \textsuperscript{2}Purdue University

“Fake news” refers to news stories fabricated and promoted on social media to deceive the public for ideological and/or financial gain. To understand the mechanisms underlying the false beliefs people acquire from fake news, we conducted online studies examining the influence of picture-text correspondence and familiarity due to prior experience with the fake news. In Experiment 1, participants rated picture-text correspondence of 24 online news headlines (half real and half fake) as they are seen on Facebook. The rated correspondence level varied across news items regardless of their legitimacy. In Experiment 2, a single exposure increased participants’ perceived accuracy of news headlines. Moreover, participants showed a better discrimination for fake and real news with low correspondence but not for those with high correspondence. The results suggest that not only repetition but also correspondence – basic cognitive mechanisms – contribute to the spread of fake news online.

8:15-9:15 AM  
Session 3  
Experimental Work I

8:15  

Testing the limits of transferability

Samuel M. Harding, David Crandall, Bennett I. Bertenthal  

Indiana University

Recent advances in machine learning (ML) have led to the deployment of robust and automated decision-making systems within a diverse array of engineering applications. State-of-the-art techniques including Deep Neural Networks (DNN) extract the latent structure within high-dimensional datasets to classify novel inputs. One limitation of these models are adversarial samples, which are inputs designed to produce robust misclassification by ML systems (changing a “3” into an “8” by modifying <10% of pixels). Human decisions including classification, discrimination, and similarity judgments, are minimally affected by adversarial perturbations (Harding et al., 2018) when allowed sufficient decision time. However, when given only 50 ms, human performance appears closer to chance (Elsayed et al., 2018). This suggests that higher-level inference plays a crucial role in mitigating the deleterious effects of adversarial inputs. We explore human judgments over adversarial stimuli, as a function of e.g. presentation time, in order to inform our understanding of the cognitive processes supporting robust human decisions in this domain.

SCiP 15
Using Machine Tools to Analyze Students’ Ethics Essays

Roman Taraban, Lakshmojee Koduru, Mark LaCour, John R. Schumacher, Micah Iserman, Taleen Nalabandian, Joseph A. Martis, Rogelio Carrillo

Texas Tech University

The present research is part of an interdisciplinary project between engineering and psychology using machine tools to analyze the ethical thinking in essays submitted by students in an undergraduate engineering ethics course. The general method in this work is to train machine algorithms on a sample of student work and compare machine and human agreement on the trained samples and on new samples, as the criterion measures. In the first part of this research, IBM’s Watson Natural Language Classifier was trained to classify sentences from students’ course papers as either related or not related to ethics, and to identify the ethical theories from the course reflected in students’ positions. In the generalization tests to new samples, Watson’s classifications fit human classifications in the range of 70-80%. An analysis of the specific topics in the course papers was undertaken using Latent Dirichlet Allocation (LDA) with results that were plausible based on informal inspection. Our current work focuses on a new set of ethics-related essays and aims to more rigorously assess agreement between human and LDA extracted topics. This research is set out to better understand how both humans and machines process language.

Validating Instagram Likes as a Measure of Aesthetic Liking

Katja Thömmes, Ronald Hübner

University of Konstanz

Be it on Facebook, Twitter, or Instagram, “Like” buttons are all over social media, generating huge amounts of data every day. In this project, we are developing methods to collect liking and engagement data from the photo-sharing platform Instagram with the purpose of validating such data as a measure for aesthetic preference in the field of empirical aesthetics. With over 1 billion active users worldwide as of June 2018, Instagram is by far the largest photo-sharing platform. In the present study, a database of 16,544 photographs from the fields of architecture, landscape, portrait, and travel (posted on 12 different Instagram accounts) was used to investigate how Likes emerge as a function of time and account size. Furthermore, we conducted a series of experiments to test how closely Instagram Likes are related to more common psychological variables, such as aesthetic liking and aesthetic preference, collected under controlled experimental conditions. In a series of online experiments, we used a three-point Likert scale as well as a visual analogue scale. Our results show that Instagram Likes are significantly correlated with the experimental liking data. Accordingly, this approach provides a promising methodological basis for future research on aesthetic principles. The validation of Instagram Likes as a measure of aesthetic liking prepares the ground for isolating image properties that might explain ‘why’ some photographs are preferred over others.
That Emoji Means What?
Exploring the Semantics of Emoji

Kay L. Livesay, Megan Kozak Williams

Linfield College

Created in Japan in the late 1990s (Davis & Edberg, 2015), emoji have recently been gaining popularity in text communication. Emoji evolved from emoticons (e.g., ; ) ), but are far more varied and nuanced. Researchers argue that emoji (like emoticons) emphasize affect (e.g. humor, sadness, sarcasm) in a text communication, thus facilitating deeper understanding of the communication by eliciting the activation of nonverbal information (e.g, Wang, 2016). It cannot be assumed that emoji will have the same effect given the diversity and continued growth of the emoji language. Understanding emoji semantics is an important first step in understanding the influence emoji have on text communication. Our research focuses on the semantic meaning of emoji in isolation and will be followed by the influence these emoji have in context. In the pilot study, 100 participants (38% male, 62% female) interpreted 42 popular emoji. Participants first described the emoji and then briefly stated what the emoji expressed to them. Based on these open-ended data we created a follow-up experiment containing a multiple-choice survey asking new participants to choose which of five meanings best expressed a particular emoji. Additionally, we asked participants to rate how positive or negative they perceived the emoji to be. The results suggest, that like words, emoji show a fan-like effect, where some have very specific semantic representations (e.g., 😄) with little variation in interpretation, while others are more ambiguous (e.g., 😑), garnering many possible meanings. Furthermore, these semantic representations may be moderated by age and gender.

9:30-11:00 AM  Session 4  Applied Computation

Chair: Randy Jamieson, University of Manitoba

Digital Traces of Identity in Online Discourse:
Detecting Socio-Cognitive Roles with Group Communication Analysis

Nia Dowell¹, Christopher Brooks¹, Oleksandra Poquet²

¹University of Michigan, ²National University of Singapore

The psychological science field is in need of new automated methodological approaches that offer deeper insights into the dynamics of group interaction and discourse across scaled online platforms. In this paper, we explore Massive Open Online Course (MOOC) learners’ discourse by employing Group Communication Analysis (GCA), a methodology for quantifying and characterizing the discourse between learners in online interactions. Commonly used approaches in MOOCs derive insight into the learning processes from aggregated text or structural data. In contrast, GCA makes use of linguistic cohesion analysis across sequences of learners’ interactions in multi-party communication. GCA calculates six inter- and intra-
personal sociocognitive measures of such interactions and from these identify distinct interaction profiles through a cluster analysis. With this method, we were able to diagnostically reveal robust profiles amongst MOOC learners. This study presents a unique analysis of the sociocognitive processes that comprise the interaction between learners. The scalability of the methodology opens the door for future research efforts directed towards understanding and improving scaled peer-interactions.

9:45

**The lifecycle of open-source software:**

**Mining GitHub to understand community dynamics**

Alexandra Paxton¹, Nelle Varoquaux², R. Stuart Geiger², Chris Holdgraf²

¹University of Connecticut, ²University of California, Berkeley

Naturally occurring data—that is, data that were not initially collected for scientific purposes—can provide a window into human cognitive and behavioral processes outside the lab. Today, whole communities can thrive nearly entirely online in open-access forums, leaving a record of community dynamics that researchers can use to understand human social behavior in real-world settings. One such forum is GitHub (https://www.github.com), a code-sharing platform used by many open-source software developers to coordinate the creation and maintenance of software. Because open-source software projects are often maintained by communities of volunteers working to sustain shared (and often vital) software infrastructure, the ability of these communities to attract and maintain new members is vital; otherwise, these projects would languish as existing community members leave. Here, we analyze community members’ interactions on GitHub to understand the social dynamics that make communities more welcoming—or more hostile—to newcomers.

10:00

**Simulating Dynamic Task Redistribution within Human-Machine Teams**

**Using an Autonomous Manager**

Mary E. Frame¹, Jennifer S. Lopez², Alan S. Boydstun¹

¹Wright State University, ²Air Force Research Laboratory

With recent advancements in automation technology, military surveillance environments must increasingly incorporate and effectively leverage autonomous systems. Concurrently, mission environments must accommodate a single agent effectively juggling multiple tasks, introducing additional workload and fatigue. Fortunately, analysts are able to offload some of these tasks to automation, but the degree of assistance autonomous systems can provide is still somewhat limited for cognitively complex tasks, requiring intelligent distribution of tasking between humans and autonomy. To address this problem of maintaining performance in a human-machine team, we developed a program that dynamically distributes tasks between human analysts and automated agents using intelligent decision logic. By re-conceptualizing the team dynamic within the current working environment, we developed a simulated supervisory role called the Autonomous Manager (AM). The AM dynamically reallocates tasks based on physiological data and task performance. This data is used to make intelligent task redistribution decisions between humans and their automated teammates. Our initial
simulation efforts provide dual benefit: 1) as an initial step toward human-in-the-loop development, and 2) as a means to save time and resources when testing multiple scenarios. For example, a new task may need to be added to a Human-Machine Team’s (HMT’s) existing workload, with multiple proposed tasks. Via simulation, the AM is capable of modeling the expected performance of the human and machine given this new additional workload. The simulated AM can be leveraged to answer a variety of real-world questions without the expense of physical implementation, prior to full-scale development and empirical testing.

10:15

**RotCorr and MDStat:**

**New ways to optimize rotation of multidimensional scaling solutions**

K. Jake Patten, Michael K. McBeath

Arizona State University

We describe a new analytic methodology that we developed to better interpret mental representation models derived from classic multidimensional scaling (MDS) solutions. MDS is a valuable tool for graphically representing the organizational structure of perceptual stimuli. MDS produces an n-dimensional plot in which distance represents the dissimilarity between stimulus objects. The overall orientation and meaning of the dimensions, however, must be determined by the researcher. RotCorr is a program that correlates a secondary, unidimensional set of data with a single axis of the MDS solution. It progressively rotates the MDS solution to determine the largest correlation between it and the secondary data set. RotCorr can identify which of several possible continua characteristics best describes the MDS dimensions, and can be used to determine the extent of orthogonality between tested secondary dimensions. MDStat is a related tool that determines 95% confidence intervals around MDS data points to identify perceptual differences between participant groups. Whereas programs like Indscal identify individual differences in group-level data, MDStat determines differences between predetermined groups. MDStat can also be used for theory testing by comparing solution locations of data points to predicted locations. We provide several research examples of RotCorr and MDStat being used to identify MDS dimensions: (1) Consonance dimension in sound perception ($r=0.78$), (2) Timbre color dimension in perception of $f_0$-matched vowel phonemes ($r=0.90$), and (3) Density dimension in weight perception of size-weight illusion stimuli ($r=0.65$). RotCorr and MDStat are handy new tools we developed to help simplify and standardize interpretation of MDS solutions.

10:30

**Using semantic memory theory to build an applied cognitive search engine**

Randall K. Jamieson¹, Harinder Aujla², Matthew J. C. Crump³, Matthew T. Cook¹

¹University of Manitoba, ²University of Winnipeg, ³Brooklyn College

We present a semantic search engine that leverages semantic memory theory to match documents by meaning. We apply the tool to documents from the psychological record to illustrate its value. The method offers a psychologically informed method for database representation, search, and retrieval.
Knowledge Structures for Medication Side Effects: Human vs. Computational Methods

Ruth S. Day, Jiehan Zheng

Duke University

Understanding possible side effects of prescription medications can help patients take appropriate action when they occur. Even if they read information about a prescribed medication, their underlying knowledge structures may affect their ability to understand and use the information accurately. To what extent do the knowledge structures of laypersons resemble those of experts? Human knowledge structures for side effects were assessed using a computer-based sorting task. Participants saw a list of side effects and sorted them into piles based on perceived “similarity” -- they could define similarity in any way they wished. The results were examined in various ways, such as hierarchical cluster trees. Although experts can perform the same task, which experts should participate – doctors, pharmacists, drug developers, or regulators who approve drugs? In order to tap the general knowledge of thousands of experts from a variety of relevant occupations, we focused on drug labels – the detailed document that must be approved by FDA regulators in order for a drug to be marketed and prescribed. We obtained 38,000 FDA-approved labels and used a variety of data mining and statistical methods to extract knowledge structures that may reflect the combined knowledge of the experts who produced them. The results show the degree of match in the overall knowledge structures of laypersons vs. computer-simulated experts. Specific discrepancies suggest potential sources of layperson misunderstanding and ways to address them. This research is a potential way to study knowledge representation across experts in multiple disciplines, without having to test them directly.

9:30-11:00 AM  Session 5  Language II

Chair: Brendan Johns, University at Buffalo

9:30

Writing for Coh-Metrix: A Systematic Approach to Revising Texts to Foster Gist Inferences

Mitchell E. Dandignac, Christopher R. Wolfe

Miami University

Flesch-Kincaid Grade Level and psycholinguistic text analysis variables measured by Coh-Metrix are proximal rather than causal measures of readability and other constructs. Thus editing texts to adhere to those proximal variables does not guarantee that texts will actually have the desired effects on readers. Following Fuzzy-Trace Theory, we used Coh-Metrix to identify psycholinguistic variables that correspond to producing coherent “gist” mental representations from text. The goal of this work is to produce a text evaluation formula that accounts for the nuance of how we form gist inferences about texts that are not captured by traditional formulas. We calculated the average of seven adjusted Coh-Metrix variables to
create Gist Inference Scores (GIS), a measure designed to assess the potential for readers to form gist inferences from observable text characteristics. Measures of referential cohesion, deep cohesion, and abstract verb overlap increase GIS because they facilitate coherent mental representations. Word concreteness, specificity, and imageability decrease GIS because they promote verbatim representations. No current literature has developed systematic procedures for altering text to adjust the level of these variables. We suggest top-down and bottom-up approaches to revising texts using Coh-Metrix variables that are more likely to affect readers’ cognitive representations. We focus on changing content words, verbs, connectives, and syntactic features within the text and make note of how different psycholinguistic variables share overlap in what textual characteristics they measure. This work has implications for experimental research by offering systematic ways to develop text stimuli. We also consider practical applications such as medical communication.

9:45

The effects of time and place on lexical behavior:
A distributional analysis

Brendan T. Johns¹, Randall K. Jamieson²

¹University at Buffalo, ²University of Manitoba

Recent research in computational psychology has demonstrated large-scale differences in the use of language across individuals (Johns & Jamieson, 2018). This presentation aims to understand the effects of time and culture on individual language usage. Specifically, this work will determine if models of language trained on a time- and culture-specific corpus of language can explain language behavior by people from that time and culture better than people from a different time and culture. To accomplish this, a large corpus of books was organized by author date of birth and place of birth. In total, the corpus includes over 25,000 books from over 3,000 authors. It will be shown that place and time of birth results in a unique signature of language usage, and that using time- and culture-specific corpora provides a better accounting of lexical behavior, giving insights into the effects of experience on language processing and organization.

10:00

Deep Learning of the basic structures of Chinese characters

Xiaowei Zhao

Emmanuel College

In this study, the printing forms (different fonts) of about 3000 common Chinese characters were sent into a Deep Neural Network (DNN), along with their sounds. It can be seen that the network can successfully learn the association between the form and the sound of these characters. The network also develops certain level of generalization when facing new characters. In addition, the internal representations on different layers of the network show the emergence of basic writing structures of Chinese characters (i.e. strokes, radicals, left-right, top-down structures …). The learning pattern of the network is further compared with that of the elementary school students.
**Using large text corpora to compute word emotionality scores: Construction and validation of a corpus-derived Chinese word emotionality database**

Yen-Cheng Chen¹, Meng-Ning Tsai¹, Yao-Ting Sung¹, Yu-Lin Chang¹, Shu-Yen Lin¹, Shu-Ling Cho²

¹National Taiwan Normal University, ²Fu Jen Catholic University

We used emotional seed terms and very large text corpora along with a multitude of association measures to automatically build a Chinese word emotionality database. A word’s emotional scores were constructed by computing the collective association strengths of the word with the seed terms. Unlike many other word emotionality databases that characterize words in only two or three dimensions (e.g., SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity), our database quantifies the emotional meaning of Chinese words on seven different scales – happiness, love, sadness, anger, surprise, disgust, and fear. To validate the effectiveness of our big-data-based approach, we compared our database with C-LIWC for their accuracy to discriminate three kinds of emotions (happiness, sadness, anger) identified by human raters of 5,230 texts consisting of 200 to 400 Chinese words. C-LIWC is the Chinese equivalent for the computerized text analysis program LIWC that outputs the percentage of words in a given text that fall into different linguistic and psychological categories. Predicting the emotional themes of the texts based on C-LIWC achieved the precision rate of 0.60. When using our database, the precision rate increased to 0.72 (p < .001). The results indicate that building an emotional database with big data has high potential for discovering the psychological meaning of words.

**Structure of scale-free random networks**

Peizhu Qian, Zhiqiang Cai, Xiangen Hu

*University of Memphis*

In this research, we attempt to analyze the structural characteristics of real-life networks by constructing simulations using distributional semantic representations for natural language. We demonstrate that there are three similarities between human networks and distributional semantic representations of language “token” networks. First, there are distinguished identities in both systems: tokens and persons, which become the nodes in their networks. Second, the nodes are linked to other nodes with a varying connection strength. In language networks, without the loss of generality, we utilize latent semantic analysis (LSA) to calculate an association strength, ranging from 0.00 to 1.00, between any two words to indicate the similarity level. In human networks, relationships can be evaluated by degrees of familiarity. Third, both of these networks are scale-free random networks, meaning only a few nodes in the network have high degree distribution, known as hubs. The Barabasi model has shown human network to be scale-free. We prove that language networks are scale-free because only some common words have high degree distributions, serving as the hubs, whereas rare words, which account for 90% of the corpus, have low centrality. Thus, we build the map between language networks and human networks. In analyzing language networks, we examine the characteristics of any subgroups, then extend them to human networks. For example, we apply
attributes of a cluster formed by financial terms to a human financial network. We use the words’ association strength to quantify the interpersonal connection strength. In addition, we calculate the average path length between any two words to further bridge language networks to human networks.

Recognition memory for the structure and meaning of text in adult readers with high and low levels of literacy.

Anne Lippert, Jessica Gatewood, Ying Fang, Genghu Shi, Qinyu Cheng, Carl Cai, Shi Feng, Su Chen, Arthur Graesser

University of Memphis

Compared to adults with high literacy levels, reading strategies of low literacy adults are suboptimal when text requires deeper levels of comprehension. To better elucidate comprehension components lacking in lower literacy populations, we tested recognition memory for the meaning and structure of text in University of Memphis students (n=153) and Mechanical Turk workers (n=70). Recognition for the surface structure of a sentence means recognizing the wording and order of the words in which a sentence was originally presented, even if the meaning of the sentence has changed. Recognition for meaning entails recognizing the semantic/conceptual message conveyed by the sentence, even if the sentence structure is changed. Since recognition of surface structure is aligned with lower level language skills and recognition of meaning requires higher level discourse components, recognition performance can help us better pinpoint comprehension deficits. Participants completed an online AutoTutor reading lesson and were assessed on recognition ability using a four-alternative forced-choice (4AFC) test. This test asked participants to identify the sentence that appeared in the lesson. The correct choice and three distractor items were displayed: text with the same meaning but different structure than the actual text; text with the same structure but different meaning than the actual text; and text that was different both in structure and meaning from the actual text. Recognition for meaning(structure) occurred when participants recognized the actual text or the text with the meaning (structure) intact. We are investigating several research questions in this study. Does literacy level indicate differences in recognition memory for surface structure versus meaning of text? Is recognition for the surface structure and/or meaning affected by aspects of AutoTutor lesson text, such as whether the verbatim sentence in the passage was a main idea versus supporting content or whether the sentence was directly queried by the conversational agents of AutoTutor.
Acquiescence in Internet-based research: An artifact of socially desirable responding?

Vaka Vésteinsdóttir\textsuperscript{1}, Ragnhildur L. Asgeirsdottir\textsuperscript{2}, Fanney Thorsdottir\textsuperscript{2}, Ulf-Dietrich Reips\textsuperscript{1}

\textsuperscript{1}University of Konstanz, \textsuperscript{2}University of Iceland

Acquiescence responding (i.e. “yea-saying” tendency, AR) and socially desirable responding (SDR) have received considerable attention. However, less attention has been paid to how the estimation of the former might be influenced by the latter. The purpose of this study was to evaluate the role of SDR in a measure of AR online, using items from the Big Five Inventory (BFI). Previous research has identified 16 item pairs with opposite implications (Soto & John, 2009). Agreeing with both items in an item pair is taken to imply AR. However, the desirability of an item’s content also plays a role in the likelihood of item endorsement. Therefore, if both items in an item pair have desirable content, the likelihood of agreeing with both items is increased, and consequently, the type of responding that would be taken to indicate AR. In this study, item content desirability (SDSV: Social Desirability Scale Value; Edwards, 1957) was evaluated for each of the 32 items belonging to the item pairs in a sample of 214 university students. The sum of the SDSV’s for each item pair was then correlated with the percentage of respondents who agreed with both items in a separate sample of 895 university students. Results showed a substantial correlation between item pairs’ total SDSV and the percentage of estimated AR, indicating an inflation in the measure of AR when both items have desirable content. Further evaluation of methods and implications of the research are discussed.

Guidelines for improving the reliability of Google Ngram studies: Evidence from religious terms

Nadja Younes, Ulf-Dietrich Reips

University of Konstanz

Google Ngram is a search engine that charts word frequencies from a large corpus of books and thereby allows to examine cultural development. While the tool’s massive corpus of data (about 8 million books or 6% of all books ever published in several widely used languages) has been exploited in various scientific studies, concerns about the accuracy of results have simultaneously emerged. This paper proposes a state-of-the-art guideline for Google Ngram studies by suggesting six hands-on approaches that can increase the reliability of results. In
particular, we recommend the use of different language corpora, synonyms, word inflations, content-free words, cross-checks on different corpora from the same language, and the consideration of standardization procedures. We further highlight that potential censorship and propaganda have to be taken into account. As an example of the proposed approaches, we examine the cross-cultural development of religion via religious terms for the years 1900 to 2000. Special emphasis is placed on the situation during World War II. In line with the strand of literature that emphasizes the decline of collectivistic values over time, our results suggest an overall decrease of religious language. However, religion re-gains importance during times of crisis such as WWII. By comparing the results obtained through the different methods, we illustrate that a consideration of our suggested approaches can increase the reliability of results. Thereby, the combination of different approaches constitutes additional robustness of the methodology.

~10:00

**Improving data quality in Internet-based studies: CRT item generator**

Deborah Canales¹, Pablo Garaizar², Ulf-Dietrich Reips¹

¹University of Konstanz, ²University of Deusto

Ensuring data quality in Internet-based research studies is a constant challenge for researchers. During the study design phase, one has to consider ways of reducing the influence of factors like social desirability, satisficing, acquiescence, attention, and familiarity on response behavior. Seriousness checks, instructional manipulation checks and “trick” questions have been introduced not only to detect how reflective a participant is while responding to a questionnaire, but even as a cue to trigger less heuristic and more analytic processes, which could in turn improve the quality of the data being collected. The Cognitive Reflection Test (CRT) has long been a popular measure of analytic versus heuristic processing and – more generally – an attention check. Several studies have included it for the purpose of selecting better quality data. However, the test’s popularity has resulted in participants becoming familiar with it. Although disputed, studies have yielded that familiarity with the CRT raised participant scores on the CRT significantly. As a measure against familiarity, researchers have contributed new items, which are isomorphic to the original three CRT items, thus increasing the item pool available for new studies. Nevertheless, these new items risk undergoing the same fate as the original CRT items. To solve the issue, we introduce CRT item generator, an online tool that generates original CRT-type items and variations of traditional CRT and attention check items.

~10:15

**Methods for studying sensitive topics online**

Ulf-Dietrich Reips, Deborah Canales

University of Konstanz

One major advantage of Web-based research lies in the anonymity of the experimental or survey situation, the Internet is thus particularly well suited for surveys on sensitive topics. For example, Mangan and Reips (2007) reported findings on how sexsomnia, a rare medical condition whose sufferers engage in sexual behavior during their sleep, could be investigated
much more easily with Web-based surveys than in all 20 years of research in that area before. However, given the online setting unconstrained by an experimenter, it is also often observed that non-response is much more frequent online. These changes may mean that sensitive topics and sensitive items ought to be handled differently when researched online versus offline. With the present paper we summarize and review methods that have been used online on studying sensitive topics. These include (1) Randomized Response technique and its variants, (2) Indirect Questioning Techniques, (3) Blurring, (4) Verification via metadata and paradata, (5) Unmatched Count Technique, (6) Methods using touch screen technology, (7) Use of non-response. Guidelines and recommendations are derived and discussed, e.g., we show how non-response measures combined with paradata predict quality of data.

Smartphone sensor data in experience sampling: The gyroscope and its heterogeneities

Tim Kuhlmann1, Pablo Garaizar2, Ulf-Dietrich Reips1

1University of Konstanz, 2University of Deusto

Smartphones are becoming increasingly popular as assessment devices in experience sampling studies. They offer an inexpensive and powerful tool to conduct longitudinal assessments in the behavioral and social sciences. In addition to self-report data, numerous sensors integrated in today's devices also offer the opportunity to gather sensor data without additional effort from the participant. We investigate the gyroscope sensor, which measures the tilt of the smartphone across its axes. Previous studies indicated that data gathered from the gyroscope sensor show promise as an indicator of body posture. They also revealed inconsistencies in the data, depending on the implemented hardware and software. Analyses showed differences between iOS and Android systems with regard to the distribution of tilt measures and the average tilt (t=2.58, p<.001, d = .12). These finding were based on post-hoc analyses from applied experience sampling studies. Thus, in the present study we examined the gyroscope sensor in a more systematic and controlled manner. Gyroscope data were assessed repeatedly from the most common smartphone models and operating systems. To ascertain the tilt, the smartphones were fixated and the tilt measured precisely via a digital protractor. In addition to the measurement of different smartphone models, we also compared data obtained via different software, e.g. different applications and browsers. Heterogeneities resulting from the different data sources were analyzed and their influences on the measurement compared. The heterogeneities are potentially of profound importance in studies that use sensor data to trigger events or to detect participant behavior like standing or lying down.

Digitally nudged: How websites influence donation behavior

Markus Weinmann

University of Liechtenstein

Nudging has attracted a great deal of attention in recent years. It is a concept for behavioral change in which psychological context effects are used instead of monetary incentives in order to actively shape peoples decision-making environment, however, without restricting any
choice. With internet-based systems being primary digital decision environments, they are particularly suitable for 'digital nudging.' In this paper, we investigate to what extent different types of nudges—especially those that trigger either intuitive thinking (Type 1 nudges) or analytical thinking (Type 2 nudges)—can be implemented in digital environments. Due to our research focus on digital environments, we conducted a completely web-based experiment (e.g., Reips, 2002), using modern methods of data collection. In particular, we used Prolific, a tool specifically designed to recruit participants for scientific studies, and distributed all study material online. Using the example of online donations, we show that nudges have a significant influence on users' donation behavior. With our results we not only contribute to the growing literature on internet-based research but also provide direct implications for practical applications.

**Discussant: Chris Wolfe**

<table>
<thead>
<tr>
<th>11:15 AM-Noon</th>
<th>Session 7</th>
<th>Experimental work II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair: Caitlin Mills, <em>University of New Hampshire</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11:15

**An improved assessment model for AutoTutor answers**

Keli Cheng, Zhiqiang Cai, Art Graesser

*University of Memphis*

With Expectation-Misconception-Tailored (EMT) conversations, AutoTutor plays an important role in developing intelligent tutoring systems. In each conversation, AutoTutor starts with a main question, which usually requires a 5-10 sentences long answer. AutoTutor helps learners construct the answer by asking a sequence of follow-up questions. There are two types of the follow-up questions. One is called “hint”, which targets an answer of a single sentence long. The other is called “prompt”, which targets an answer of a word or phrase. The performance of AutoTutor relies on accurately assessing learner’s answers to AutoTutor questions. This paper reports an improved model for assessing student answers to AutoTutor hints and prompts. A dataset collected from a physics AutoTutor application was used, which contained 4941 student responses to 114 hint questions and 2643 responses to 133 prompt questions. All responses were rated by two human experts in a 0-6 scale. The dataset was split into four subsets: hint training/testing set and prompt training/testing set. We applied the ridge regression model provided by Md Arafat Sultan et al. (2016) to this dataset and got results similar to what they reported, with correlation $R=0.422$ for hints and $R=0.660$ for prompts. By adding a domain-specific semantic to the model, the correlation is improved to $R=0.510$ for hints and $R=0.670$ for prompts. A neural network model based on these features further improved the accuracy to $R=0.530$ for hints and $R=0.710$ for prompts.
A real-time mind wandering intervention during reading

Caitlin Mills¹, Julie Gregg², Robert Bixler³, Sidney D'Mello²

¹University of New Hampshire, ²University of Colorado Boulder, ³University of Notre Dame

Our minds wander off-task approximately 20-40% of the time during reading—a rate that is concerning given the negative relationship between mind-wandering (defined here as off-task thought) and comprehension. Mind-wandering is thought to attenuate processing at basic levels of encoding during reading, leading to poor information integration at the conceptual level. Here we tested if a real-time learning intervention during mind-wandering episodes could mitigate its negative impact on reading comprehension. When an eye-gaze-based detector predicted that a reader was currently mind-wandering, participants in a mind-wandering-sensitive condition (MW) were asked to self-explain and potentially re-read parts of the text in order to improve their conceptual understanding of what they just read. Each participant in the MW condition was paired with a yoked-control counterpart (YC) who received the exact same interventions regardless of their mind wandering. Surface- and inference-level comprehension were assessed immediately after reading, then again after a week. Comprehension scores were compared beginning from the time each pair received their first intervention since no differences were expected up to that point. The two conditions performed equally well on both the surface- and inference-level comprehension questions immediately after reading, p>.05. After a week-long delay, however, the MW condition (M=.449; SD=.200) significantly outperformed the YC condition (M=.389; SD=.191) on inference-level comprehension questions, B=.293, p=.040, d=.307, and marginally significantly on surface-level questions (MW: M=.585; SD=.176 vs. YC: M=.531; SD=.176), B=.192, p=.056, d=.309. These findings suggest that real-time interventions during critical periods of mind-wandering can promote better long-term conceptual comprehension.

Computer- and paper-based normative ratings of metaphors:
They are similar in some ways and different in other ways

Cesar E. Riano Rincon, Gary E. Raney

University of Illinois at Chicago

The purpose of our study is to compare metaphor norms collected using computers compared to norms collected by paper. We conducted a norming study using metaphors of the “X is a Y” format (e.g., history is a mirror) taken from Campbell and Raney (2015). Participants rated (1) how familiar the metaphors were, (2) how many interpretations they could think of, and (3) how related were the two content words. One group responded using computers and one group responded on paper, but all ratings were collected in person (not online). Ratings collected by computer and paper were very highly correlated (r > .97 for all three ratings). However, the paper group provided significantly higher absolute ratings of familiarity and relatedness, and significantly lower number of interpretations than the computer group. For example, for the paper group, the average familiarity rating for the 20 least familiar metaphors was actually higher than the average for the 20 mid-familiarity metaphors from the paper group. We discuss how standardizing ratings can minimize this problem. In addition, more participants produced
“bad” data (e.g., responding 1 to every question) in the computer group than the paper group. This might cast doubt on the quality of computer-based norms. Whether you choose paper or computer normed stimuli might not seem important given that the ratings were highly correlated across methods. Even so, we conclude that care must be taken when using norms that are collected using different methods as covariates in analyses or when matching stimulus properties.

1:00-2:00 PM

Poster Session

[Posters are ordered by ascending order of first author’s last name.]

Poster# 147

A general, modular toolkit for semantic model implementation and research

Johnathan E. Avery, Michael N. Jones

Indiana University

Numerous toolkits exist for the generation of semantic spaces. Popular models such as GenSim and TensorFlow are highly streamlined software packages that allow users to extract semantic similarities from a given corpus. However, these packages do not employ many of the most successful semantic models, leaving a wide swath of parameter space in semantic research under explored. For instance, GenSim employs continuous-bag-of-words (CBOW) and skip-gram, as well as other statistical techniques, such as LSA, LDA, and td-idf transformations, while TensorFlow is limited to the CBOW and skip-gram models. We present a generalized, modular toolkit that implements often overlooked models of semantic memory. These models include HAL, COALS, GLoVe, and BEAGLE. Additionally, the toolkit fosters an exploration of model combination and construction allowing for an effective framework for future research into the efficacy of techniques in semantic modeling.

148

ERTS Scripting Language for Browser-based Reaction Time Experiments

Joerg Beringer

Berisoft Inc.

This paper introduces a re-launch of the ERTS Scripting Language for defining web-based reaction time experiments. ERTS was a popular DOS-based software package offering a simple scripting language to define reaction time tasks. Unlike source code generators like EPrime or code libraries like jsPsych and psychoPy, the scripting language is shielding the researcher from technical aspects of programming and managing source code. ERTS Scripts are very descriptive and are directly interpreted by a Javascript runtime environment that can execute the script in any modern web browser. This architecture makes it the ideal tool for sharing and publishing cognitive task paradigms. In combination with the Cognition Lab web platform, ERTS scripts can be deployed as browser-based test applications with data being automatically collected.
Promotion of complex password generation through privacy priming

Isis Chong, Robert W. Proctor, Ninghui Li, Jeremiah Blocki

Purdue University

Password security and usability are often at odds as increased complexity makes passwords less memorable while simple passwords are easily hacked. Although many users have a general idea about what makes a password secure, they often report using the same password or cycling through similar passwords for the sake of ease and habit. Past research by the authors has suggested that privacy priming can be an effective tool in reconfiguring a user’s task set to more heavily weight security-related information in decision making for a simulated app selection task. Given the potential for other cybersecurity domains, the present study was aimed at investigating whether priming can be generalized to promote more complex password string generation. We utilized a simulated password generation task with participants recruited through Amazon Mechanical Turk. Participants created a password on a page with either (1) neutral, security-unrelated, (2) general, security-related, or (3) specific, security-related material. Security-related information was solely centered around password management and did not provide information on how to complete the generation task. Password complexity was assessed through the ratio of alphanumeric to non-alphanumeric characters, overall length of passwords, and other factors. The results suggest that privacy priming can indeed be generalized across domains to promote privacy-focused behaviors. Password complexity was found to vary as a function of the priming presented during generation. Recommendations for improving cybersecurity-related behaviors through relatively simple interventions based on these findings will be discussed.

Applying the feature model to immediate serial recognition

Chrissy M. Chubala, Ian Neath, Aimée M. Surprenant

Memorial University of Newfoundland

In an immediate serial recall task, participants are asked to recall a short list of items in their original presentation order. In an immediate serial recognition task, participants are given a second list containing the same items either in the same order or with two items transposed and are asked to make a same/different judgement. Although empirical work has explored similarities and differences between the two tasks, no model has attempted to explain both. We examine the extent to which the Feature Model (Nairne, 1990), a model of serial recall, can be extended to explain serial recognition. In the model, items are represented as vectors of features, and cues become degraded through interference. At test, the degraded cues are compared to intact items in secondary memory. In serial recall, the item with the highest relative match is selected for output. We incorporate a version of Farrell & McLaughlin's (2007) model of serial recognition, in which the overall similarity between the two lists is computed to determine a response. Within the Feature Model, this is instantiated using the mean number of mismatching features. The combined model adequately captures many aspects of performance in serial recognition and makes new predictions, whereas its failures suggest alternate avenues for exploration.
Sequence effects in dating: Comparisons between Tinder and speed-dating

Samantha Cohen

Indiana University

Mobile dating is one of the most popular methods for meeting new romantic partners and offers a rich, high external-validity method for studying mate choice. Previous researchers studying romantic decisions “in the wild” have turned to speed dating, a face-to-face sequential mate-choice paradigm. How do these two methods of mate choice compare? To contrast these two areas, I characterize another sequential mate-choice paradigm: Tinder. Despite its immense popularity, the little work on Tinder has focused solely on motivations of use, rather than choice behavior. In this work, I recorded the displays of 38 consenting undergraduate Tinder users in the lab as they were presented with partners in the app. I then make comparisons to a commercial speed-dating dataset from 16 events and over 400 users. In both Tinder and speed-dating, daters are presented with partners in sequence, with a decision (like/dislike) made after each partner before encountering the next. I examined effects of order between the two paradigms. Neither group exhibited any sequence effects in terms of varied rates of choosiness over the course of the experiment (primacy or recency). However, Tinder users exhibited a significant assimilation effect; after liking a profile, both men and women were much more likely to like the next profile as well (the effect also presents for disliking behavior). In speed-dating, previous work has demonstrated contrast effects. I discuss why behavior may vary between these two paradigms and the implications for the mate choice field as we transition into studying mates in this new context.

Using REM.3’s superimposition-of-similar-images assumption to model spacing and list-strength effects

Tyler M. Ensor, Aimée M. Surprenant, Ian Neath

Memorial University

First described by Shiffrin and Steyvers (1997), the Retrieving Effectively from Memory (REM) model has successfully accounted for many empirical phenomena in the memory literature. REM incorporates a differentiation assumption, wherein strengthened items are accumulated in a single mnemonic trace rather than in separate traces. Differentiation allows REM to account for several benchmark findings, including the null list-strength effect in recognition (Ratcliff, Clark, & Shiffrin, 1990). To date, instantiations of REM have largely treated different strengthening methods identically—that is, spaced and massed repetitions and increased study time have been modeled in the same way, with a single trace being updated. However, Shiffrin and Steyvers’ original description of REM discussed the possibility of a superimposition-of-similar-images assumption. According to this assumption, repetitions are Only included in a single trace if the subject identifies that the repeated item was previously studied. Here, we show that incorporating this assumption allows REM to predict three important findings in the literature: the spacing effect, the finding of slightly positive list-strength effects with spaced repetitions as opposed to massed repetitions or increased study
time, and list-strength effects that have been observed using very large strong-to-weak ratios (see Norman, 2002). This latter finding in particular has been used as evidence against the viability of REM. Critically, though, when considering this more realistic version of REM, that criticism is no longer valid as the model predicts the observed pattern of performance.

GazeR: A package to analyze gaze position and pupil size data

Jason Geller¹, Matthew B. Winn², Tristan Mahr³, Daniel Mirman¹

¹University of Alabama at Birmingham, ²University of Minnesota, ³University of Wisconsin-Madison

Today, eye-tracking is widely used throughout the scientific community, from visual science and psycholinguistics, to marketing and human-computer interaction. However, most laboratories use in-house scripts for data processing and analysis, resulting in surprisingly little consistency and making replicability difficult. To increase replicability and transparency, a package in R (a free and widely used statistical programming environment) called gazeR (https://github.com/dmirman/gazer) was created to read in and preprocess two types of data from the SR EyeLink eye tracker: gaze position and pupil size data. For gaze position data, gazeR has functions for: reading in raw eye-tracking data, formatting it for analysis, converting from gaze coordinates to areas of interest, binning and aggregating data, and various helper functions for multilevel regression (growth curve analysis) and plotting. For data from pupillometry studies, the gazeR package has functions for reading in and merging multiple raw pupil files, eliminating missing data that exceed a specified threshold, linear interpolation of blinks, binning and aggregating data, and subtractive baseline correction. We present data from two tasks exemplifying the package’s capabilities.

Beyond the lab: Collecting mouse-tracking data in online studies

Felix Henninger¹, Pascal Kieslich²

¹University of Koblenz-Landau, ²University of Mannheim

Cognitive scientists working in the laboratory are increasingly including mouse-tracking in their studies to assess the development of cognitive processes over time, particularly the emergence of choices and the degree of conflict between response options. At the same time, an increasing amount of experiments are conducted online, allowing for efficient data collection in larger and more diverse samples. The present contribution addresses both of these developments, and explores the possibility and unique challenges of collecting and analyzing mouse-tracking data in self-administered online studies, where the conditions of data collection are not as easily controlled. Based on these, we introduce a platform-independent, open-source JavaScript library (https://github.com/FelixHenninger/mousetrap-web) that can be integrated in existing web-based studies to collect and analyze mouse movement data.
Functional ANOVA for analyzing response dynamics

Joseph W. Houpt\textsuperscript{1}, Mary E. Frame\textsuperscript{2}

\textsuperscript{1}Wright State University, \textsuperscript{2}Wright State Research Institute

Dynamic responses have been leveraged increasingly over the last ten years to analyze various cognitive processes. These measures, particularly mouse trajectories, have yielded important insights into topics ranging from word perception to decision-making. Along with the increase in use, a number of statistics have arisen for comparing trajectories. These include summary statistics, such as the area between the observed trajectory and direct path from start point to choice location (AUC), and the frequency of directional changes along the choice axis (x-flips). While these statistics have proven valuable, they filter out much of the information about the trajectory itself and remove all of the temporal information. In some papers, sequences of t-tests are used to capture some of the temporal properties of a trajectory, however this is technically a misuse of the test. To solve this problem, we propose utilizing functional analysis of variance (fANOVA), a statistical model for examining how independent variables of interest affect functional data, such as trajectories. fANOVA can be used to analyze dynamic responses while capturing the richness of the full trajectories without the shortcomings of sequential t-tests. We will cover a re-examination of mouse trajectories from a decision-making under risk task and an examination of the differences in response trajectories elicited from three different input devices.

A comparative study of DNN and CNN for fMRI data classification

Jinlong Hu\textsuperscript{1}, Yuezhen Kuang\textsuperscript{1}, Shoubin Dong\textsuperscript{1}, Ping Li\textsuperscript{2}

\textsuperscript{1}South China University of Technology, \textsuperscript{2}Pennsylvania State University

Deep Neural Network (DNN) and Convolutional Neural Network (CNN), the two main types of deep learning architectures, have been successfully applied to analyze various functional Magnetic Resonance Imaging (fMRI) data (Kriegeskorte et al., 2015). DNN can model complex non-linear relationships and CNN has been found to excel at extracting position-invariant features. In this presentation, we experimentally compared models of DNN, 2D CNN and 3D CNN and applied them to two open fMRI datasets that involve cognitive tasks. We examined model accuracy and model training time, and used the three types of model to classify a large number of time-series whole-brain imaging data. Our preliminary findings indicate that convolution operations had advantages in high-dimensional whole-brain imaging data classification; Extracting 2D images from 3D whole-brain imaging data for 2D CNN processing may result in loss of meaningful information; 3D CNN models achieved highest accuracy in classifying the fMRI data, although these models require more computational resources than other models, and they also take longer time to train properly.
Assessing non-Euclidean spatial representations in virtual reality

Po-Han Huang, Christopher J. Widdowson, Ranxiao Frances Wang

University of Illinois at Urbana-Champaign

In Euclidean space the distance between parallel lines remains constant into infinity. However, there are other non-Euclidean spaces, such as 3D spherical space (S^3), where parallel lines intersect, and 3D hyperbolic space (H^3), where parallel lines diverge. Previous work has implemented visualization techniques for 3D non-Euclidean spaces but is ultimately limited to a desktop or tablet style interface. The applications are also limited by shading and projection techniques that cannot offer sufficient depth cues to communicate non-Euclidean distance in a truly immersive way. The present work further develops the techniques from Weeks (2016) and ports this configuration to a modified virtual reality (VR) interface. The application reads tracking information from the HTC Vive VR display and hand-held controllers for view-projection and navigation, respectively. Stereo-specific modifications (e.g. field of view, interpupillary distance (IPD), and motion scale) were made to accommodate a non-Euclidean metric space. A distance-based fog effect is included that adjusts the brightness of objects, proportional to the inverse square of the distance to the user. The resulting simulation depicts a non-Euclidean 3D universe with earth-like planets of different colors floating sparsely in space. The users turn their body and move a virtual beam pointer using a hand-held controller to the direction they want to go and press a button to travel forward along the geodesic of the curved space. An application has been developed for a navigation experiment to examine people’s ability to complete a point-to-origin task for two non-Euclidean environments (S^3 and H^3) and a Euclidean control.

The SCiL application - Conducting community-driven cognitive research

Felipe de Luca¹, Adrian Cunningham², Curt Burgess², Mabel Wong², Bruno Flora¹, Kimberly Miller²

¹Centro Universitário Senac, ²University of California, Riverside

At last year’s meeting, we presented the Semantic Cognition in Language (SCiL) project, a collaboration between the Computational Cognition Lab, University of California, Riverside, an Apple-sponsored developer team at the Centro Universitário Senac in São Paulo, Brazil. The demonstration of this app showed the app’s significant potential in validating the lexical decision methodology. The goal of this project has been and continues to be to develop an application for iOS and Mac devices that can be used to conduct cognitive experiments using reaction time, thus providing for experimental methodologies for research of memory and language function. Consistent with this goal, SCiL is developed to be HIPAA and GDPR compliant to provide reliable security for sensitive user data. Using SCiL, researchers are able to set up cognitive experiments. Participants can download the SCiL app from the App Store and browse available experiments. Implemented procedures include the lexical decision priming task, the categorization task and a questionnaire module. The demo now includes a downloadable brief semantic priming experiment for demo purposes and researchers can...
implement original experiments. In this presentation, we introduce the SCiL Builder web app (https://builder.myscil.org/) where researchers can implement priming experiments or simple decision-making protocols and make them available for research participants to download and participate in these experiments. This is the next step of our long-term goal to develop a world-wide community of researchers who can use the site for collaborative projects that will capitalize on this technology.

Web based Q-methodology: The app for that

Susan Lutfallah, Lori Buchanan

University of Windsor

Quantitative analysis of qualitative data can be accomplished using Q-Methodology. This research method has been used to study participants’ subjective views on various topics. The task has historically been completed by manually sorting cards into categories that force responses into a normal distribution. We will describe a computerized web-based Q-Sort application that allows multiple participants to sort their responses at the same time, in any location, negating the need for research studies to be attended by researchers, and results to be manually recorded. The program provides automatic data collection for immediate examination and can be exported in a consistent manner for external data analysis. This program has been used in both applied and cognitive studies. For example, in a clinical study it is being used to evaluate participants’ perception of behaviours seen as most characteristic or most uncharacteristic of psychological aggression or coercive control in situations of intimate partner violence. In a psycholinguistics lab it was used to evaluate the affective valence, abstractness, and semantic richness ratings of words. We examined how the Q-Sort ratings of valence and abstractness mapped onto pre-existing, computer generated ratings for those variables, to determine whether any assumptions could be made about the semantic richness ratings of those words.

Predicting language ability from eye movements during reading:
Utilization of Deep Neural Network (DNN)

Xiaochuan Lindsey Ma¹, Lijie Cao², Ping Li¹

¹Pennsylvania State University, ²South China University of Technology

Eye movements have been utilized as an index of attention and comprehension during reading from a wide range of literature. Highly skilled readers are found to show shorter fixations, more skips, and fewer regressions as compared with less skilled readers (Ashby, Rayner, & Clifton, 2005). In this study, native English speakers (L1) and second-language English learners (L2) read five science texts, while their eye movements were recorded. We were interested in identifying whether eye-movement features can serve as reliable indices for readers’ language ability, and hypothesized that these two groups of participants with different levels of proficiency in English could be classified given their eye movement features (word-level fixation duration, skip rate, and regression rate, etc.) during reading. A Deep Neural Network (DNN) model (Bengio, 2009) with two hidden layers was trained on the eye-
movement features to learn to predict the two groups (L1 vs. L2 participants). Different parameters of the DNN model were evaluated in randomly selected training data and test data. After applying L1 & L2 regularization as well as dropout layer to reduce over-fitting, the DNN model reached a 71% accuracy in classifying L1 vs. L2 readers, reflecting different information intake patterns possessed by readers with high vs. low language ability.

Implementing the outer loop in iSTART: Adapting text difficulty as feedback

Kathryn S. McCarthy¹, Amy M. Johnson², Micah Watanabe², Danielle S. McNamara²

¹Georgia State University, ²Arizona State University

Interactive Strategy Training for Active Reading and Thinking (iSTART) is an intelligent tutoring system aimed at improving reading comprehension through self-explanation. iSTART uses natural language processing to evaluate students’ self-explanations (0-3). Currently, the system provides “inner loop” feedback in the form of targeted messages that offer suggestions on how to improve the self-explanation. This study explored the potential of adding “outer loop” feedback to iSTART. Outer loop feedback adapts the difficulty of the task based on students’ previous performance. We present the development and implementation of an adaptive logic that uses students’ average self-explanation score to increase or decrease text difficulty. High school students engaged in 7 hours of iSTART practice in which they read complex science texts and received points for generating high quality self-explanations. All students received inner loop feedback messages. In the adaptive condition, an average self-explanation score above the threshold (2.0) increased the difficulty of the subsequent text, whereas a self-explanation score below this threshold decreased the text difficulty. In the control condition, students received texts in random order. Preliminary results indicate that adapting text difficulty may be particularly beneficial for low-skilled readers, both in terms of in-system performance and post-training learning outcomes.

How do word frequency and number of related senses interact in different tasks? Insights from dictionary data mining, and mega-study analyses

Di Mo, Barend Beekhuizen, Suzanne Stevenson, Blair C. Armstrong

University of Toronto

Most words are polysemous, denoting multiple related but distinct senses (‘chicken’ as ANIMAL/FOOD/COWARD). Understanding how degree of polysemy interacts with word frequency is therefore key to developing models of lexical processing. Most studies have not looked for such interactions, reporting facilitation for more-polysemous words in lexical decision and inhibition for more-polysemous words in semantic categorization. However, Jager, Green, and Clelland [JGC] (2016, LCN) reported a three-way interaction between frequency, polysemy, and task, showing worse performance for high-frequency, high-polysemy words in lexical decision and for low-frequency, high-polysemy words in semantic categorization. We used computational methods on large-scale data to better understand the frequency-polysemy relationship. We extracted sense definitions from the Wordsmyth

SCiP 36
Dictionary, deriving various measures of polysemy: raw counts of dictionary senses, and more nuanced counts that reflect relatedness of senses (using dimensionality-reduction applied to distributional vector representations of definition words). We tested for interactions between frequency, each polysemy measure, and task on polysemes from the British Lexicon Project and the Calgary Semantic Decision Project (an order of magnitude more data than previous studies), and on the items from JGC. In lexical decision, we found an advantage for more-polysemous words that was attenuated at high frequencies. In semantic categorization, we found a disadvantage for more-polysemous words that did not interact with frequency. While our relatedness-corrected sense counts produced only minimally improved fits, their correlations with raw counts reveal that dictionary definitions have consistent sense-granularity across words. Our results reinforce the benefit of large-scale data methods in informing models of lexical processing.

How do I meet everyone in the world during my lifetime?

A social network simulation with NLP techniques

Peizhu Qian, Zhiqiang Cai, Xiangen Hu

University of Memphis

This research utilized natural language processing (NLP) techniques to simulate a human society to answer a social network question: how many acquaintances does a person need in order to meet everyone in the world? In 1998, based on six degrees of separation, Watts and Strogatz suggested that a person needed to have 30 acquaintances to connect to 90% of the world population. However, their formula had limitations since people shared mutual friends in real life so the minimal number of acquaintances must be larger. With current conditions, it is almost impossible to obtain this number by conducting a real-life experiment. Therefore, this research constructed a simulation in which a word (node) represented a person, and the corpus formed the population. As people were connected to each other by familiarity, the words were associated together by a similarity value (edge) calculated by latent semantic analysis (LSA). The LSA method analyzed the context in which two words were appeared, and utilized the k-nearest neighbor (kNN) algorithm to determine the semantic similarity. In addition, to better understand the linguistic interaction, the authors utilized Gephi, a network analysis software, to perform statistical analysis and visualize the expanding network. The results presented a cogent mapping between language association and human social network. This study took the initiative to practice NLP tools in social science and demonstrated how computing simplified currently unfeasible experiments. Building upon existing papers, this research took more realistic and practical considerations and made contributions to an interdisciplinary subject.
Effects of gamification on an online platform

J. Antonio Salamanca, David A. Washburn

*Georgia State University*

Tasks commonly used by cognitive psychologists require effortful, sometimes costly expenditure of cognitive resources. These tasks can be described as fatiguing or even boring to participants. Recently, there has been an increased interest gamification, or, the adoption of game-like techniques in non-game tasks (see Lumsden, Edwards, Lawrence, Coyle, & Munafò, 2016). We explored the effects of gamification on common measures of executive control. Preliminary results (n = 50, data collection ongoing) suggest that participants are faster to respond on “go” trials of a go-no-go task when the instructions contextualize the task as a game (game condition mean = 441ms, nongame condition mean = 466ms, t = 2.22, p = .03). Additionally, participants who received the game-like instructions showed higher error rates on no-go trials (i.e., greater difficulty inhibiting a response, game condition mean = 21%, nongame condition mean = 7%, t = 3.76, p < .001). Participants also completed a psychophysical discrimination task with an uncertain response option. Participants in the game-instruction condition show more adaptive responsiveness to uncertainty when compared to the nongame-instruction condition (means = 26% and 23% respectively). Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *JMIR Serious Games*, 4(2), e11. [http://doi.org/10.2196/games.5888](http://doi.org/10.2196/games.5888)

Open Lab: A web application for conducting and sharing online-experiments

Yury Shevchenko¹, Felix Henninger²

¹University of Mannheim, ²University of Koblenz-Landau

Online experiments have become a popular way of collection data in social sciences. However, high technical hurdles in setting up a server prevent a researcher from starting an online experiment. On the other hand, proprietary software restricts the researcher’s freedom to customize or share the study. We present Open Lab – the server-side application that makes online data collection simple and flexible. Available online at [https://open-lab.online](https://open-lab.online), the application offers a fast, secure and transparent way to deploy the study. It takes care of uploading experiment scripts, changing test parameters, managing the participants’ database and aggregating the study results. Open Lab is integrated with the lab.js experiment builder ([https://lab.js.org/](https://lab.js.org/)), which enables the creation of new studies from scratch or the use of templates. Saved as a JSON file, the study can be uploaded to Open Lab and is ready to run. At the conference, we will present the main features of Open Lab together with results of the pilot study conducted at the University of Mannheim.
Getting in Touch: Method and technical considerations for studying tactile temporal-order judgments

Ty Y. Tang, Michael K. McBeath
Arizona State University

We describe methodology and instrumentation for performing high-precision temporal-order judgment experiments. In these experiments, precise temporal resolution and synchronization measurements are critical to capturing the temporal relationship between events. In our implementation, we (1) Compare the timing of participant responses to onset-times of a variety of sensory stimuli, and (2) Compare the timing differences between two participants’ responses. In such experiments, confounds such as clocking-asynchrony, buffer-lag, and other hardware and software constraints can bias and corrupt measurements of temporal relationships. The design we created eliminates most hardware issues by having all peripheral stimuli and response equipment interfaced with a single central processor clock. We also mitigate software timing issues by carefully designing programmatic loop structures to minimize processing time and Boolean redundancy. Finally, we made considerations for the software programming structure when designing our experiment to further reduce the impact of software runtime. The sensor equipment for measuring temporal differences includes wearable capacitive sensors for capturing touch timing, along with LEDs and fast-acting solenoids for generating stimuli. The sensor equipment input/output voltages are directly interfaced with an Arduino microprocessor kit, which logs the timing data and sends it upstream to a Windows PC for processing. In addition to presenting the apparatus described above, we will also discuss considerations for capturing auditory responses and muscle electrical activity, and we will review various implementations and limitations of the design. Our approach provides a simple template for building and executing accurate multisensory timing experiments.

A joystick for every monkey, redux

David A. Washburn, J. Antonio Salamanca, Will Whitham, Mattea S. Rossettie
Georgia State University

Consumer access to highly efficient, low-cost single-board computers (e.g., Raspberry Pis (RPi)) have been a boon for STEM educators and hobbyists alike. We explored the viability of using such devices in another application - as a potential replacement to the legacy hardware that is widely used in cognitive-behavioral research. Specifically, we attempted to use RPi hardware and software to accomplish the litany of tasks required in computer-based cognitive testing of nonhuman primates in a research-lab setting. We demonstrated that RPis are exceedingly well-suited to such an application. That computers and monkeys are shared by multiple labs from multiple research and programming traditions demands support for many legacy hardware and software families. RPi is able to accommodate this constraint via rich package and emulation support, as in emulation of DOS-based devices for software written in TurboPascal and emulation of x86 processor architectures for software written as Windows
executables. Required hardware peripherals - including joysticks and dispensers of appetitive rewards to the primates - are also supported. The small size of the RPi enables more options for protecting sensitive hardware components from the threats of heat and moisture posed by standard nonhuman primate housing. RPis' low cost and highly replaceability, especially in comparison to relatively-higher cost and problematic reliability of legacy hardware options, is a further boon. We present a summary of RPi viability in a specific comparative cognition research application, and a more general description of the advantages and disadvantages of their use in psychological testing.

Learning outcomes for student-generated E-tutorials vs other content presentation modes

David A. Washburn

Georgia State University

In teaching the History of Psychology, it is a challenge to help students connect personally with the material. The learning outcomes from several classroom activities designed to address this challenge are reviewed in this poster. In particular, an e-tutorial activity will be compared to student presentations, traditional lecture, and textbook-only modes of information presentation. For e-tutorials, students created 10-minute lessons on a significant and under-discussed person, theory, laboratory, or place in the history of psychology. The lessons could be narrated Powerpoint or Prezi presentations, PowToon videos, TED-style talks, or any other audiovisual format, so long as the tutorial was self-executing, visually engaging, and factually accurate. Most etutorials were distributed as short movies via YouTube. Students’ familiarity with and knowledge about the individuals encountered in these e-tutorials were compared to changes in familiarity and knowledge regarding historical figures/developments encountered from (a) student-led academic ancestry presentations, (b) traditional lecture by an instructor, or (c) textbook or supplemental reading only. A start-of-term pretest established baseline levels of knowledge for each student, and the end-of-term posttest included items from each of these four presentation methods, plus items that were not covered in any way during the semester, as a control for effects of repeated testing. Familiarity ratings and fact recall showed improvements for all four methods of presentation, with readings-only significantly worse than the other three options. E-tutorial presentations showed positive learning outcomes and highly positive student evaluations. Additionally, of course, they yield instructional content that can be used in subsequent semesters.
Domain-general cognition in another's domain: Methodological considerations for computer-based statistical learning research with nonhuman primates

Will Whitham, David A. Washburn

Georgia State University

Statistical learning is a domain general cognitive mechanism that operates over many different kinds of stimuli to synthesize meaningful structure from disparate individual items, as when children learn the structure of their native language from repeated, unrewarded exposures to its constituent items. Nonhuman animals, too, have demonstrated the ability to learn the probabilistic structure embedded in stimuli, as when rhesus macaques (Macaca mulatta) evidence faster responding for repeated sequences in a Serial Reaction Time task (Heimbauer, et al., 2012) or golden lion tamarins (Saguinus oedipus) learn the sequential order of list items (Locurto, Gagne, & Nutile, 2010). As a complement to these successes, we present evidence of the relative fragility of statistical learning phenomena in computer-based testing of nonhuman primates. Across multiple experimental manipulations (e.g., to stimuli, probabilistic structure, or task parameters), empirically-validated statistical learning tasks (e.g., Serial Reaction Time, probabilistic categorization), and species (rhesus macaques and tufted capuchin monkeys (Cebus apella)) we demonstrate only inconsistent evidence of this domain-general cognitive capacity. We will review these results with reference to established findings from humans, and describe how the interactions among species differences in perceptual and cognitive faculties and accommodations made when translating computer-based cognitive tasks from humans to nonhuman primates can work to slow application of statistical learning principles to nonhuman animal cognition.

Relationships between personality, online behaviors, and the performance in the Iowa gambling task on Facebook

Lee-Xieng Yang, Yi-Ting Yu

National Chengchi University

In this study, we conducted the traditional Iowa gambling task (IGT), administered the Big-Five personality test, and collected the digital records (e.g., like number, share number, frequency of tagging friends in my post, etc.) for 234 participants (with their consent) with our APP on Facebook. After testing, each participant’s performance in the IGT was fit by the Value-Plus-Perseveration RL model. With 8 best-fit parameter values and the task score as the IGT data, the 5 scores on the Big-Five personality test as the personality measure, and 7 kinds of digital records as the online behavior, we examined with SEM the relationships between personality, online behavior, and performance in the IGT. The results suggest a Personality variable composed of the traits of Openness, Extraversion, and Agreeableness. Also, the task score, the parameter estimate of loss aversion, the probabilities of staying with the same option on the next trial (respectively for the positive and negative current score), and the shape of utility function are suggested to link to the IGT variable representing the IGT performance. The digital records cannot be integrated as any variable. Thus, we examine them individually.
The mediator analysis with Personality as the mediator shows that the frequency of tagging friends in my post is positively correlated with Personality, which is negatively correlated with IGT. That is, the more often a person tags friends in her/his post, the more open, extravert, and agreeable s/he is and the less likely s/he wins in the IGT also.
for millisecond accurate responses, sound, and device communication