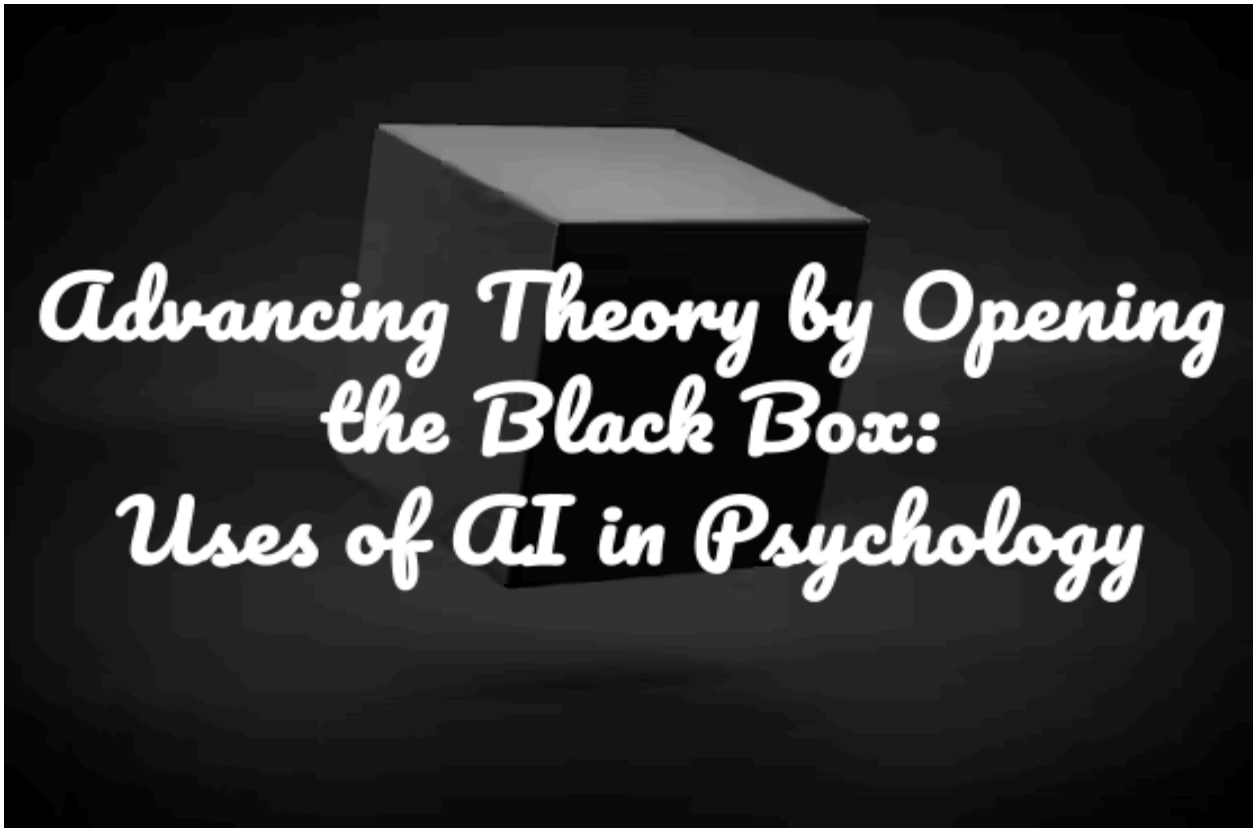


53rd Annual Meeting of the Society for Computation in Psychology

November 16th, 2023

Table of Contents:

SCiP
Society for Computation in Psychology
since 1971

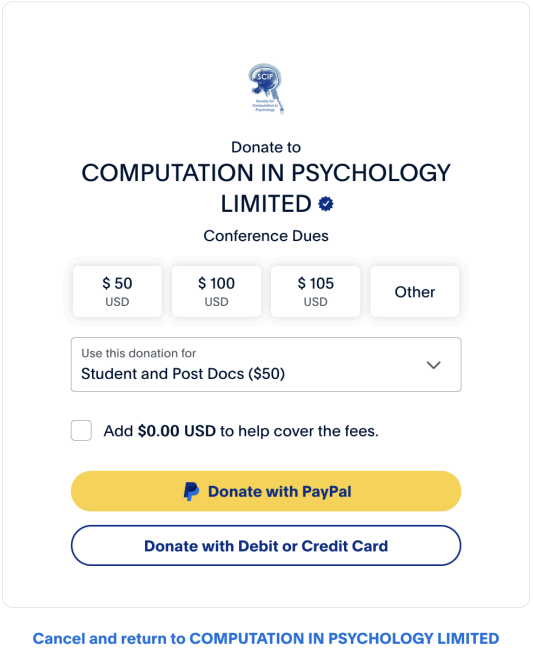
Pay Your Dues/Conference Fee

Note: Official Receipts will be sent after the conference. Please email compinpsych@gmail.com if you need a special version. Check your spam if you do not see a receipt by 11/30/23.

You can pay dues on site with US dollars or a check at the conference, please see amounts below.

Use this link to pay your dues!

- Pick a price point below:
  - Student and Post Docs: $75.00
  - Faculty/Independent Researchers (people with jobs who've graduated): $125.00
  - I feel like covering some extra fees: $130.00
- If your university/job is covering the cost of the conference, consider covering the fees for us! (not necessary, just an option).
- If you use a paypal account, we will see the name associated with that account. If you pay directly with a credit card (no PayPal account necessary), your information will be grabbed from that page. If you are paying for someone else, please leave a note or send an email to compinpsych@gmail.com so we can make name tags.
- We are registered 503(c), which makes your donation a tax exemption if you are in the United States! This registration also lowers the credit card/paypal fees.



You can also use this QR code:

Plan Your Presentation

Preparing Your Talk

Presenter and Co-author Names

The opening slide should include the names and affiliations of all the contributing scientists. The name and affiliation of the presenter should be clearly indicated.

Illustrations

Figures should be designed to be viewed from a variety of screens, using clear, visible graphics. Although each figure should illustrate no more than one or two major points, figures need not be simple. The main points should be clear without extended viewing, but detail can be included for the knowledgeable viewer.

Each figure or table should have a heading of one or two lines in large type stating the "take-home" message. Detailed information should be in smaller type in a legend below the graphic. Because there is no text accompanying a poster, the figure legend should contain commentary that would normally appear in the text of a manuscript (results and discussion). It should describe concisely not only the content of the figure but also the conclusions derived from it. Details of methodology should be kept brief and should be placed at the end of the legend.

Statistics

Please follow the Psychonomic Society Statistical Guidelines. Effect sizes should be reported and error bars with appropriate labels should be included on all graphs.

Layout

Arrange materials in columns rather than in rows. It is easier to scan a poster by moving systematically along it rather than by zigzagging back and forth in front of it. An introduction should be placed at the upper left and a conclusion at the lower right, both in large type. The sequence of illustrations should be indicated with numbers or letters at least 1-inch high, preferably in bold print.

You may find it convenient to have a separate section describing methods, but it is quite effective to include this information as part of the data presentation, as described above. Carefully chosen photographs of apparatus, or schematic diagrams of procedures, can convey a great deal of information about methods without much text. Most viewers will tend to skim or ignore long textual passages.

Day of Presentation

All presentation times are in U.S. Pacific Time, the local time for the on-site meeting in San Francisco.

Please note: as far as we understand it, you will give your presentation to a tech person at the back of the room before the session. Extra time has been allotted for this procedure this year, please load your talk before your session.

Keeping SCiP Time

Please note the difference times for presentations marked for each session. In general, they are 12 minutes plus 3 minutes for Q&A.

Session chairs and speakers should only start a talk at the scheduled time. If talks begin before the scheduled time, attendees may miss your presentation. Keeping SCiP Time also means ending on time. If a speaker doesn't finish on time, session chairs should ask the speaker to conclude the presentation so the next talk can begin at the scheduled time.

Presenting On-Site in San Francisco: The Day of the Your Presentation

- Meet your Moderator (aka Session Chair) a few minutes before the session starts in the meeting room.
- When speaking, make sure to face the microphone for good-quality sound.
- Venue staff will be nearby for assistance.
- Give your presentation to the local tech person.

Presenting On-Site in San Francisco: Audio-Visual Equipment

Each session room will have a standard audio-visual equipment, including:

- A projector, screen, and remote control
- If your presentation requires use of your personal laptop, please make sure that your laptop has an HDMI connection. The venue may not have other connectors available.
- It is **strongly recommended** that you bring a back-up copy of your final presentation to the session on a USB memory stick in case you experience any internet connectivity issues and cannot download the session from the cloud.

Poster Requirements: Printed Posters

Poster Dimensions

For proper display at the venue, your poster should have a horizontal (or landscape) orientation and be no larger than the following dimensions:

Width: 4 feet (1.22 meters)

Height: 4 feet (1.22 meters)

Title/Banner

The title of your poster should be prominent, generally a banner across the top and visible to attendees from a distance. Lettering should be at least 1-inch (2.54 cm) high.

Poster boards are numbered and ordered by the poster number included below on the schedule.

Presenter and Co-author Names

The poster should include the names and affiliations of all the contributing authors included with the abstract in the program. The name and affiliation of the presenting author should be clearly indicated.

Illustrations

Figures should be designed to be viewed from a variety of screens, using clear, visible graphics. Although each figure should illustrate no more than one or two major points, figures need not be simple. The main points should be clear without extended viewing, but detail can be included for the knowledgeable viewer.

Statistics

Please follow the Psychonomic Society Statistical Guidelines. Effect sizes should be reported and error bars with appropriate labels should be included on all graphs.

Layout

Arrange materials in columns rather than in rows. It is easier to scan a poster by moving systematically along it rather than by zigzagging back and forth in front of it. An introduction should be placed at the upper left and a conclusion at the lower right, both in large type. The sequence of illustrations should be indicated with numbers or letters at least 1-inch high, preferably in bold print.

You may find it convenient to have a separate section describing methods, but it is quite effective to include this information as part of the data presentation, as described above. Carefully chosen photographs of apparatus, or schematic diagrams of procedures, can convey a great deal of information about methods without much text. Most viewers will tend to skim or ignore long textual passages.

Conference Locations

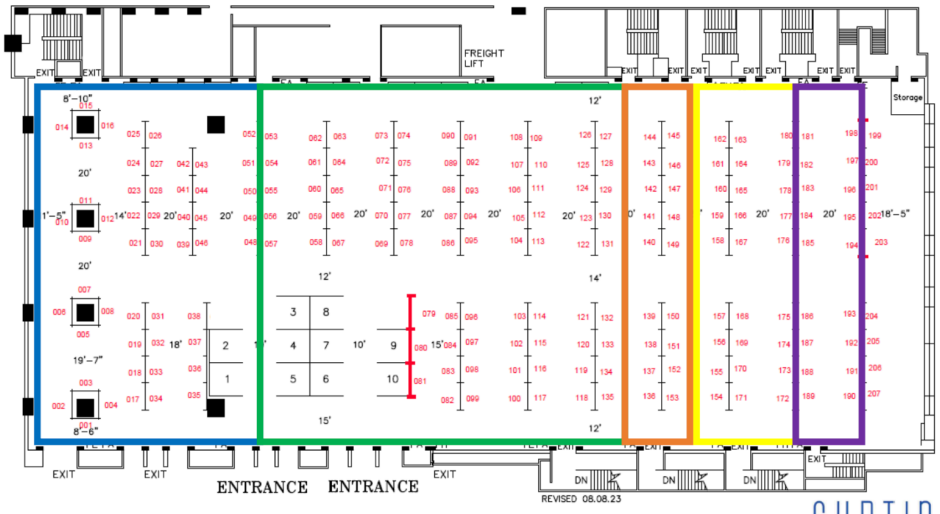
Hilton San Francisco Union Square

South Lounge, Ballroom Level (BR)

Continental Ballroom Rooms 1-2 and 3

APCAM – 50 posters requested / 51 provided
OPAM – 80 posters requested / 81 provided
SCIP – 10 posters requested / 18 provided
SIDIC – 25 posters requested / 27 provided
SMP – 10 posters requested / 18 provided

Psychonomic Society 2023 Conference
November 15–17, 2023
San Friancisco Hilton & Towers
Grand Ballroom
San Francisco, California



NOTE:
- (90) DOUBLE–SIDED POSTER BOARDS
- (30) SINGLE–SIDED POSTER BOARDS
- (10) 10x10 BOOTHS

Schedule of Presentations

All times San Francisco, CA, Pacific Time.

Continental 1-2		Continental 3	
8:00-9:00am	**Session 1:** Emotion	8:00-9:00am	**Session 2:** Methods
9:15-10:30am	**Session 3:** Large Language Models	9:15-10:15am	**Session 4:** Software
11:00-1:00pm	**Session 5:** Machine Learning	10:30-12:30pm	**Session 6:** Symposium: Internet Based Methods
Poster Session 12:30-1:30pm Grand Ballroom (GB), Access from Tower 2 in Hotel Lobby			
1:30-2:45pm	**Session 7:** Language	1:15-2:45pm	**Session 8:** Thinking
3:00-4:00pm	**Presidential Symposium** Continental 1-2		
4:00-5:00pm	**Keynote Speaker** Continental 1-2		
5:00-5:30pm	Business Meeting Continental 1-2		

Note. A fuzzy-trace set of labels to sessions. See abstracts below.

SCiP Sponsors

Our thanks to our sponsors:

[Psychonomic Society]

[Psychology Software Tools]

[Cognition Lab]

Session 1 (8:00-9:00 AM): Emotion

table of contents quick schedule

Session:	Emotion
Time:	8:00-9:00 AM (*12 minutes + 3 minutes Q&A)*
Room:	Continental 1-2
Chair:	Laura Allen
Talk 1:	***How Do You Feel and Why?: Integrating Affective and Motivational Research with a 2-Stage Self-Reporting Tool*** Stephen Hutt, University of Denver, stephen.hutt@du.edu, Jaclyn Ocumpaugh, University of Pennsylvania, jlocumpaugh@gmail.com, Nidhi Naisar, University of Pennsylvania, nidhinasiar@gmail.com This paper presents the development and implementation of a novel two-stage affect survey, which facilitates in-the-moment self-reporting of both student's emotions and the underlying reasons for those emotions. The first stage of this tool is aligned with the academic emotions outlined by Pekrun (2006, 2008), including nervous, bored, frustrated, focused, happy, and confused. The second stage extends the inquiry, offering choices aligned with constructs from motivation theory such as challenge/difficulty, eureka, utility, interest, self-efficacy, and external control (Linnenbrink-Garcia and Patall, 2016, Sinatra, Heddy and Lombardi, 2015). By distinguishing, for example, boredom stemming from high or low difficulty from boredom due to perceived lack of utility, the survey provides more nuanced insights into the student experience. The survey's design was iteratively refined with both domain experts and students to ensure that it remained consistent with theoretical principles while also using language that was suitable and understandable for the learners' age group. The implementation has been field-tested in an online learning environment. Preliminary results demonstrate that the two-stage design can successfully elicit rich and context-specific responses from students, paving the way for more personalized and effective instructional interventions. This research contributes to the broader understanding of student emotions in online learning and offers a practical tool for educators and researchers seeking to monitor and respond to students' affective states in real-time. In doing so, it also lays the groundwork for more nuanced research into the relationship between academic emotions, motivation, and online learning.
Talk 2:	***The Influence of Writing Instructions on Linguistic Inquiry and Word Count (LIWC) Category Counts*** Aleksei Proskurin, Texas Tech University, aleksei.proskurin@ttu.edu, Dr. Roman Taraban, Texas Tech University, roman.taraban@ttu.edu This study investigates the impact of different writing instructions on the scores generated by the Linguistic Inquiry and Word Count (LIWC; Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, 2022), a program that formed basis for AI sentiment analysis, as a preliminary step to establishing a neural basis for LIWC scores. LIWC analyzes language samples to provide insights into emotional and cognitive states. The present study aims to assess differences observed in LIWC scores due to variations in writing instructions provided to participants. A total of 172 participants, all native English speakers, were randomly assigned to either a happy or unhappy writing condition. The participants' LIWC scores were analyzed based on their written responses to specific prompts. The results

	revealed significant differences in LIWC positive and negative emotion scores between participants recalling happy and unhappy memories. The study highlights the ability to develop tasks that elicit targeted LIWC categories to ensure reliable and consistent results. These findings emphasize the need for carefully selected writing instructions in future LIWC research. By identifying the impact of writing conditions on LIWC category distributions, this study establishes an experimental paradigm for subsequent experiments seeking to establish a neural basis for LIWC outcomes, using FNIRS or related technologies. The establishment of the neural basis for LIWC outcomes serves as a preliminary step towards achieving a deeper understanding of how AI can leverage neuroscientific insights to enhance its emotional intelligence capabilities and create more empathetic interactions with users. This experiment opens the door to more sophisticated applications, such as emotionally intelligent chatbots and virtual assistants.
Talk 3:	***Avoiding Positivity at All Costs: Evidence of Reward Devaluation in Educational Contexts*** Mya Urena, University of Minnesota Twin Cities, urena014@umn.edu, Samuel Winer, The New School, winere@newschool.edu, Caitlin Mills, University of Minnesota Twin Cities, cmills@umn.edu Reward devaluation theory (RDT) posits that some depressed individuals may respond negatively to positive material (i.e., devaluing reward), going so far as to actively avoid it. Although there are intuitive everyday life consequences for individuals who “devalue reward” or positivity, limited research has established if (and how) reward devaluation manifests in more ecological tasks. The current research assessed if such devaluation presents in a novel Valence Selection Task in three experiments. In all three experiments, participants engaged in an activity where they read incomplete reading prompts and were instructed to choose from a positively-valenced, negatively-valenced, or neutral sentence ending. Experiments 1 and 2 findings indicated that individuals who reported a higher fear of happiness (a key symptom of depression associated with reward devaluation) were less likely to choose positive endings (E1: rho = -0.31, p = 0.005; E2: rho = -0.38, p<0.001). Experiment 3 replicated these same findings (E3: rho = -0.23, p = .023) despite the positively-valenced answer being the most correct. Overall, participants were less likely to choose the positively-valenced sentence endings, providing support for RDT in a novel Valence Selection Task with educational implications. Furthermore, individuals who reported greater fear of happiness tended to select negatively-valenced responses, consistent with existing depression literature. No significant relationship was observed between individuals' fear of happiness and their inclination to select neutral sentence endings, demonstrating that these findings were not due to non-valenced-based impairment. The findings underscore the relevance of RDT in real-world contexts and highlight potential educational applications.
Talk 4:	**Context F**king Matters: An Investigation into the Dynamic Factors of Swearing in Movie Dialogue** Lauren E. Flynn, University of Minnesota, flynn598@umn.edu, Laura K. Allen, University of Minnesota, lallen@umn.edu Taboo language in varied discourse contexts has typically elicited strong claims about the speakers of such language, such as their personalities, intelligence, and politeness. However, research rarely examines the context in which such language occurs, despite the fact that speakers’ language is strongly tied to explicit environmental contexts. We therefore argue that research on such language should not occur in isolation of its surrounding context. The current study addresses these contextual issues by leveraging publicly-available movie transcripts (n=580) to investigate: 1) how taboo language such as swearing fluctuates across time (over course of movie) with regard to contextual

	factors such as the genders of the speakers/listeners, 2) how these fluctuations relate to the dynamics of detected sentiment, and 3) how movie context (genre) influences these patterns. All transcripts were organized into sequential dyadic conversations and merged with available metadata, including speaker (e.g., gender, character name), conversation/utterance (e.g., both character names, movie name, utterances), and movie (e.g., genres, release year, IMDB rating) levels of information. A previously-published dictionary of taboo words and their affective norms were used to categorize taboo words into 3 discrete categories based on their level of perceived tabooness. Concordance analysis was used to find each taboo word and its surrounding context (5 words to the left/right of each detected word) before additionally performing sentiment analyses on the overall, pre-swear, and post-swear utterances. Our preliminary findings highlight the necessity of accounting for dynamic social factors in utilizing NLP for emotion analysis.

Session 2 (8:00-9:00 AM): Methods

table of contents *quick schedule*

Session:	Methods
Time:	8:00-9:00 AM (*12 minutes + 3 minutes Q&A)*
Room:	Continental 3
Chair:	Caitlin Mills
Talk 1: withdrawn	***Combining Knowledge Visualization and Intelligent Tutoring to Support Learning in STEM Education: The Development of KVIS (Knowledge Visualization Intelligent System)*** *Anne Lippert, Prairie View A&M University, amlippert@pvamu.edu, Donggil Song, Texas A&M University, creative@tamu.edu* *Students pursuing Science, Technology, Engineering and Mathematics (STEM) degrees often face complex problems (e.g., engineering design, environmental planning, disease spread) for which they struggle to conceptualize valid solutions. Emerging technologies like knowledge visualization make it possible to represent the learner's conceptualization or knowledge of the problem space externally (e.g., Kim, 2019). This knowledge representation can then be compared to an expert's knowledge structure or assessed with qualitative and quantitative measures indicative of stable solutions. However, unless assisted by a human instructor, students often lack the insight to analyze how or where their representation is insufficient (Bloom, 1984; Afzal et al., 2019). The traditional "one teacher to many students" classroom model makes it unlikely students receive individualized scaffolding to maximize learning gains through knowledge visualization alone. The need for students to receive personalized instruction to navigate their knowledge structures may be solved by intelligent tutors. Intelligent tutors try to replicate the benefits of one-to-one, personalized tutoring- often using pedagogical agents that converse with students (Graesser et al. 2014). Our project combines knowledge visualization and intelligent tutoring into a single, innovative tool to support STEM learning. We are currently completing the development of KVIS (Knowledge Visualization Intelligent System) - a prototype knowledge visualization tool embedded with an artificially intelligent (AI) tutor/coach. The present talk will provide details of the design and implementation of KVIS, as well as planned studies to assess KVIS's effectiveness and areas of design improvement to support STEM learning.*
Talk 2:	Experimental Validation of a New Axiomatically Derived Scoring Rule for the Subset Selection Response Format Tim Angelike, University of Duesseldorf, tim.angelike@uni-duesseldorf.de, Birk Diedenhofen, University of Duesseldorf, birk.diedenhofen@uni-duesseldorf.de, Jochen Musch, University of Duesseldorf, jochen.musch@uni-duesseldorf.de A problem with multiple-choice (MC) knowledge tests is that random guessing introduces error variance, and partial knowledge cannot be rewarded because it is not made visible. To address these problems, Zapechelnyuk (2015) proposed a modified response format that allows test takers to choose any number k of available options, and a novel scoring rule that provides partial credit by awarding 1 / k points if the correct answer is among the k options chosen. This alternative response format and the

	associated axiomatically derived scoring rule discourage random guessing, allow and reward partial answers, and satisfy the desirable properties of being easy to implement and providing non-negative scores. To compare the two competing formats, we established a strong validation criterion by experimentally inducing different levels of knowledge in a domain of which participants had no prior knowledge. In Study 1, ZAXS scores were found to be more reliable and valid than MC scores when attempting to reclassify participants to their known level of knowledge based on the scores they obtained on the two competing scoring procedures. In Study 2, we manipulated the tendency to select few options rather than many options when uncertain about the correct answer to investigate whether ZAXS scores were contaminated by differences in participants' risk taking. We found no evidence for a moderating effect of risk-taking on the superiority of ZAXS. Taken together, our results suggest that ZAXS is a promising alternative response format and scoring method for multiple-choice knowledge tests.
Talk 3:	Stability of content of thoughts Nabil Al Nahin Ch, University of Minnesota, nabilch@umn.edu, Caitlin Mills, University of Minnesota, cmills@umn.edu, Laura Allen, University of Minnesota, lallen@umn.edu, Jolie Wormwood, University of New Hampshire, jolie.wormwood@unh.edu, Julia Kam, University of Calgary, julia.kam@ucalgary.ca Many studies examining task-unrelated and freely moving thoughts have employed ecological momentary assessment (EMA) paradigms to assess these constructs in daily life. However, there is significant variability across studies in terms of sample size, number of probes per day, duration of the study (in terms of weeks), and compliance rates– ultimately leaving many unanswered questions about best practices for EMA studies in this field. Here we conducted a month-long study to address some of these gaps. We find that people's self-reported thoughts are more stable beyond the first week of reporting. Similarly, the effect sizes between concurrent EMA dimensions are more stable beyond the first week. These findings remain consistent, irrespective of the compliance rate. Our results also demonstrated that these constructs and the relationship among them stabilize with increasing sample sizes, the number of probes per day, and the duration of the study.
Talk 4: withdrawn	*Using variable selection and mixture modeling characterize teamwork perceptions during long-duration space flight simulations* *Shane T Mueller, Michigan Technological University, shanem@mtu.edu, Elizabeth S Veinott, Michigan Technological University, eveinott@mut.edu, Kathleen Mosier, TeamScape, LLC, kmosier@sfsu.edu, Ute Fischer, Georgia Institute of Technology, ute.fischer@gatech.edu* *Researchers often use ratings of team members to understand the dynamics of small-group structures and teamwork. When the ratings are absolute assessments of teamwork competency of each team member by each other team member, standard network sociograms may not be appropriate, and other analytic approaches may need to be developed. To address this gap, we examined data sets from small team (5-6 member) missions from two ground-based, extended (4-8 month) space flight analog simulations. These missions involved weekly ratings of each crew member by all other members on the teamwork behavior they exhibited. To examine the data, we developed a 3-part regression model using variable selection techniques to separately estimate (1) overall rating biases of individual raters; (2) mean perceived competency of individuals by other team members that often evolves over time; and (3) pairwise (rater-ratee) deviations from the bias+competence model. In addition to these major factors, 10-20% of individual ratings were not well captured by the model, which we identified using finite*

	mixture modeling, identifying exceptions that might indicate particularly notable events that deviate from baseline teamwork ratings. These were primarily negative deviations from a default, presumably because of specific observations on the day of the rating. Results also reveal that individuals within a team often have very different understandings of individual teamwork competencies, even when careful instructions and rubrics are used; that substantial variability in ratings is accounted for by individual conflicts within the team, which are often symmetric.

Session 3 (9:15-10:30 AM): LLMs

table of contents *quick schedule*

Session:	Large Language Models / Deep Learning
Time:	9:15-10:45AM (*12 minutes + 3 minutes Q&A)*
Room:	Continental 1-2
Chair:	Rick Dale
Talk 1:	***"Inconceivable!": LLM Generated Text does not Exhibit Plausible, Human-like Conversational Dynamics*** Zachary Rosen, UCLA, z.p.rosen@ucla.edu Because of their ability to generate "plausibly human text" (Brown et al., 2020; Cai et al., 2023), a number of new research programs have proposed the use of Large Language Models (LLMs)–a class of transformer language models–as a plausible model or replacement for human participants in psycholinguistic experiments (Gilardi et al., 2023; Hardy et al., 2023). However, LLMs' ability to generate plausibly human-like text does not necessarily entail that they plausibly engage in human-like discursive behavior. Thus, assessing LLMs' discursive ability is vital for our understanding of the applicability of LLM generated text in answering a number of psycholinguistic and sociological questions. In the current study we used an information theoretic measurement–built over a smaller, much more heavily validated transformer model to generate semantic representations of lexical units–to test the degree to which OpenAI's flagship LLM for chat–chatGPT–replicated the same discursive behavior as reddit community members when asked to write replies to a number of comments. We find that unlike human participants, the lexico-semantic content of chatGPT's responses have statistically significant lower surprisal with the comments they reply to than human responses do. In a word: chatGPT does not exhibit the same randomness in discourse that humans utilize to drive web-based communication forward. Our results thus indicate that studies that use chatGPT as a stand-in for a conversational participant may fail to capture key attributes of human discourse.
Talk 2:	My Face and the Faces of Generative AI: Predictors of FACES Christopher R. Wolfe, Miami University, WolfeCR@MiamiOH.edu, Mackenzie M. Blazek, Miami University, blazekmm@miamioh.edu, Paige A. Renschler, Miami University, renschpa@miamioh.edu, Lauren M. Lucina, Miami University, lucinalm@miamioh.edu, Anne O. Hardy, Miami University, hardyao@miamioh.edu, Grace E. Tirzmalis, Miami University, tirzmage@miamioh.edu, Deepak G. Krishnan, University of Cincinnati Medical Center, gopaladk@ucmail.uc.edu Facial Appearance as Core Expression Scales (FACES) is designed to assess maxillofacial surgery patients' perceptions of their faces. Previous research provides evidence of reliability and validity. We conducted two studies about how self-perceptions correspond with FACES ratings of participant's own faces and those produced by the generative artificial intelligence (AI) application DALL•E. Study 1 investigated individual differences with 171 participants rating their own faces and completing the Body Image Avoidance Questionnaire (BIAQ), Rosenberg Self-Esteem, State Self Esteem Scale (SSES), the Fuzzy-Processing Preference Index (FPPI, used for quality control), and other instruments. BIAQ, Rosenberg, and SSES each predicted FACES outcomes with

	R2=0.43 indicating a strong relationship between self-esteem and people's perceptions of their faces. For Study 2, we used DALL•E to generate 16 photo-realistic faces. Images are of male and female faces aged 20, 30, 40, and 50 of differing apparent ethnicities. Four stem descriptions were used for four sets of four images, for example "The face of a 40 year old male of mixed South American and African ancestry wearing something dark in the style of a professional photo portrait." This was followed by descriptions taken from all seven FACES items, such as "the face is (not) like I want others to see me." Image generation instructions differed only in "not" being used for half of the images. Participants rate images using FACES testing the hypothesis that FACES distinguishes between faces generated by positively and negatively worded instructions. We discuss DALL•E as a source of experimental stimuli.
Talk 3:	***Transformability, Generalizability, but Limited Diffusibility: Comparing Global vs. Task-Specific Language Representations in Deep Neural Networks*** Yanru Jiang, Department of Communication, University of California Los Angeles, yanrujiang@g.ucla.edu, Rick Dale, Department of Communication, University of California Los Angeles, rdale@ucla.edu, Hongjing Lu, Department of Psychology, University of California Los Angeles, hongjing@g.ucla.edu This study investigates the integration of two prominent neural network representations into a hybrid cognitive model for solving a natural language processing (NLP) task, where pre-trained large-language models (e.g. BERT) serve as global learners and recurrent neural networks (e.g. RNN-LSTM) offer more "local" task-specific representations in the neural network. Both NLP models, BERT and LSTM, are applied to the same tweets from a tweet-emotion classification task to generate global and task-specific representations, respectively. To explore the fusion of these two types of representations, two autoencoders, BERT-to-LSTM and LSTM-to-BERT, are employed. These autoencoders assess whether these representations can be transformed into each other through reconstruction and whether they can share the same embedding space through similarity measurements. After successfully generating the six representational systems (i.e. LSTM, BERT, BERT-to-LSTM diffused and reconstructed, and LSTM-to-BERT diffused and reconstructed), they are employed in a different emotion-related task, hate speech classification, to examine which systems exhibit better transferability and generalizability. Our exploration identifies a computational constraint, which we term limited diffusibility, highlighting the limitations of hybrid systems that operate with distinct types of representation. Specifically, global representations may require a higher dimensionality with higher-entropy encoding; task-specific may usually be tuned to lower-dimensional and lower entropy formulations of a specific task. The findings from our hybrid system confirm the crucial role of global knowledge in adapting to a new learning task, as having only local knowledge greatly reduces the system's transferability.
Talk 4:	***Distributional Models on Compositional Generalization Inference*** Shufan Mao, University of Illinois Urbana-Champaign, smao9@illinois.edu, Philip Huebner, University of Illinois Urbana-Champaign, info@philhuebner.com, Jon Willits, University of Illinois Urbana-Champaign, jwillits@illinois.edu Are distributional learning mechanisms capable of complex linguistic inference? We investigated this question by comparing various distributional language models (word2vec, RNNs, GPT2, and graphical distributional models) on learning and generalization tasks, to explore the computational mechanisms capable of inferring on unseen lexical combinations. We generated an artificial corpus consisting of sentences

	like “John preserved cucumber with vinegar”, containing combinatorial dependencies such as having the instrument noun jointly dependent on two other elements of the verb phrase (the specific verb and the semantic category of the direct object (e.g., the vinegar could only occur in sentences with the verb preserved and with a direct object that was a VEGETABLE). We trained different models on the corpus, and tested whether they learned the combinatorial lexical dependencies, by their predictions for the instruments in novel verb phrases. Correct predictions could be obtained only if the models had formed a compositional representation of the sentence: representing the distributions of the words individually as well as with regard to the phrases in which they participated. In these experiments, static models like word2vec failed to learn the dependencies. While RNNs learned the dependencies from input, only GPT-2 and the CTN model (a graphical network formed by joining constituent parse trees) succeeded in generalization. We show the success of the two models was due to attentional mechanisms and explicit constituent parse tree respectively, and argue that some emerging structure in the multi-layer attention blocks may contribute to representing and generalizing complex combinatorial lexical dependencies in natural language.
Talk 5:	**Word Surprisal Predicts Language Learner Proficiency** Langdon Holmes, Vanderbilt University, langdon.holmes@vanderbilt.edu, Scott Crossley, Vanderbilt University, sacrossley@gmail.com An important component of psycholinguistic research is the role of prediction in language processing (Clark, 2013; Huettig et al., 2022). Research has demonstrated that language processing difficulty is proportional to word surprisal, an information-theoretic construct that measures the unexpectedness of a word (Futrell et al., 2020; Goodkind & Bicknell, 2018; Rommers et al., 2020). This study applies word surprisal to the context of second language acquisition. We calculate surprisal for each word in TOEFL independent essays using a large language model (LLM), which provides word probability predictions for each word in an essay. We test the extent to which word surprisal can predict the writing proficiency assigned to each essay using a support vector classifier, which ranked student essays into low, medium, and high proficiency levels. The model correctly classified 75% of learner essays in a balanced subsample using only 2 textual features: word count and word surprisal. The model represents a more parsimonious approach to classifying language proficiency that also shows improvement in accuracy over previous attempts using the same dataset (Vajjala, 2018). Results indicate that more proficient language learners use words that are more predictable according to an LLM, suggesting that predictive competency plays an important role in learner language development.

Session 4 (9:15-10:15 AM): Software

table of contents *quick schedule*

Session:	Internet-based research: Methods, tools, research synthesis
Time:	9:15-10:15 AM (*12 minutes + 3 minutes Q&A)*
Room:	Continental
Chair:	Erin M. Buchanan
Talk 1: (withdrawn)	***Power Analysis with G*Power 4.0: More Flexibility for Complex ANOVA and MANOVA Designs*** *Edgar Erdfelder, University of Mannheim, Mannheim, Germany, erdfelder@uni-mannheim.de, Franz Faul, Christian Albrechts University, Kiel, Germany, ffaul@psychologie.uni-kiel.de, Albert-Georg Lang, Heinrich Heine University, Düsseldorf, Germany, Albert.Lang@hhu.de, Axel Buchner, Heinrich Heine University, Düsseldorf, Germany, axel.buchner@hhu.de* *G*Power 4.0 is a major extension and improvement of the popular G*Power power analysis program. To facilitate use of the new version, program handling and the types of power analyses remain essentially unchanged compared to previous versions. However, some procedures (e.g., ANCOVA) have been updated to simplify insights in parameters that may affect power. Most importantly, generalized effect size measures and special procedures for repeated measures ANOVAs and MANOVAs have been added that can handle basically any factorial design, irrespective of the number of factors and factor levels that vary between or within participants. Another major addition is the General Linear Model (GLM) Explorer, a tool to calculate effect sizes and power from user-specified mean and covariance structures under H1 for basically any design. We illustrate the new features, discuss good practices of program use, and provide information about G*Power 4.0 availability (which will be free, like previous G*Power versions).*
Talk 2:	***STAPLE: Your New Favorite Science Software*** Erin M. Buchanan, Harrisburg University of Science and Technology, ebuchanan@harrisburgu.edu This talk will demonstrate a science focused project management tool called STAPLE: science tracking across the project lifespan. Scientific research has become increasingly complex, requiring specialized skills, interdisciplinary work, and collaboration among large teams. Thus, managing such projects and tracking data and metadata has become a significant challenge. Therefore, there is a need for scientific project management software that is tailored to the management of all manner of science projects while simultaneously aiding the collection and curation of scientific metadata. We will present the current project status of STAPLE, showing off current software capabilities, along with the plan for features and other integrations. Attendees (and other interested persons) will also be invited to participate in the developmental stage of the software by testing and providing information about bugs, usability, and design.

Talk 3:	***An Introduction to MultiSOCIAL Toolbox: An Open-Source Library for Quantifying Multimodal Social Interaction*** Veronica Romero, Department of Psychology and Davis Institute for Artificial Intelligence, Colby College, USA, vcromero@colby.edu, Tahiya Chowdhury, Davis Institute for Artificial Intelligence, Colby College, USA, tchowdhu@colby.edu, Alexandra Paxton, Department of Psychological Science & Center for the Ecological Study of Perception and Action, University of Connecticut, USA In everyday life, most of our experiences of social interaction are multimodal, but in our research, most of our studies of social interaction focus on only one kind of communication behavior. However, by reducing the dimensionality of our data, we are limited in our ability to capture the dynamics of real-world social behavior. A major reason for this unimodal focus is technological: Until recently, the richness of human behavior in just minutes of face-to-face interaction could take over an hour of meticulous hand-coding, transcription, and annotation, but advances in computing power and software innovation are changing that. Here, we present a new effort to assemble open-source tools into a single platform for multimodal interaction data: the MultiSOCIAL Toolbox (or the MULTImodal timeSeries Open-SourCe Interaction Analysis Library). While these tools exist in separate packages for scientists with programming abilities, our goal is to expand access to scholars with limited (or even non-existent) programming experience and to accelerate discovery through a unified multimodal data processing pipeline. The toolbox enables any researcher who has video files of any kind of interaction to extract time-series data in three modalities: body movement (non-verbal behavior); transcripts (what was said during interaction); and acoustic prosodic characteristics (how it was said).
Talk 4:	***High-Performance Experiments in R with the Glia Package*** Felix Henninger, Ludwig Maximilian University of Munich, mailbox@felixhenninger.com The R programming language is now widely used, and taught, for data processing in psychology and many social sciences. Despite its popularity and the widespread familiarity with R among students and researchers, when it comes to building experiments, both groups have had to pick up entirely new programming languages and skillsets. The glia package aims to remedy this, providing an accessible toolkit for constructing and running experiments from R. However, R was not originally designed to meet the demands of low-latency interaction and fast stimulus presentation that are fundamental to laboratory research. Therefore, we investigate whether, and how, it can be adapted, to combine a familiar and accessible programming language with the needs of scientists in terms of presentation latency and response timing accuracy. We demonstrate that the glia package can achieve high-fidelity timing from R, validate its performance with regard to stimulus presentation and response time measurement, and present a general pattern for building performant experiments in environments not built for this purpose.

Session 5 (11:00-1:00 PM): Machine Learning

table of contents *quick schedule*

Session:	Reading
Time:	11:00-1:00 PM (*12 minutes + 3 minutes Q&A)*
Room:	Continental 1-2
Chair:	Rick Dale and Stephen Hutt
Talk 1:	***A Cognitive Science Rosetta Stone for Model Interpretability: Mapping the Learning Curves of Deep Learning Networks*** Yanru Jiang, Department of Communication, UCLA, yanrujiang@g.ucla.edu, Rick Dale, Department of Communication, UCLA, rdale@ucla.edu Over the last decade, neural networks have exhibited impressive performance across prediction and classification tasks spanning diverse domains. However, the growing intricacy of deep neural networks (DNNs) poses challenges for understanding the mathematical basis of model predictions. Cognitive science and computational neuroscience have long examined internal model representations, often seeking connections to human cognitive processes. Building upon this foundation, our paper introduces an innovative technique to analyze the learning trajectories of DNNs. We term this tracking a “learning curve,” as it resembles research on measuring a human learner under different situations. This study proposes a model-interpretability method that can extract the learning curve of sequence-based deep learning networks (e.g., RNN-LSTM). Inspired by cognitive science, the method measures the learning trajectory and underlying knowledge extracted by such networks. To approximate the learning curve of a model’s confidence across the LSTM timesteps, this study used multiple traditional statistical models (e.g. logit, KNN, SVM) to examine the information retention throughout all timesteps. Each model is applied to data points (x_t, y), where x_t represents the LSTM embedding at timestep t, and y is the corresponding output label. Tracking the learning trajectory of a DNN enables us to explicitly identify the model appropriateness of a given task, while also examining the properties of the underlying input signals. To illustrate the method, we use temporal tasks, gesture detection, and natural language processing (NLP) as examples, showcasing its applicability across a spectrum of deep learning tasks.
Talk 2:	***Shallow or Deep Conversation? Understanding the Multimodal Dynamics of Interpersonal Connection*** Grace Qiyuan Miao, University of California, Los Angeles, q.miao@ucla.edu, Joyce Yanru Jiang, University of California, Los Angeles, yanrujiang@g.ucla.edu, Ashley Binnquist, University of California, Los Angeles, abinnquist@ucla.edu, Agnieszka Pluta, University of Warsaw, apluta@psych.uw.edu.pl, Francis Steen, University of California, Los Angeles, steen@comm.ucla.edu, Matthew Lieberman, University of California, Los Angeles, mdlieber99@gmail.com, Rick Dale, University of California, Los Angeles, rdale@ucla.edu

	Talking to another person is one of the most common activities that people engage in. Studies have shown that although most daily conversations consist of small talk, more substantive and meaningful conversations tend to make people happier (Mehl et al., 2010; Milek et al., 2018). While the psychological processes underlying shallow and deep conversations have been extensively examined (Kardas et al., 2022), the neurocognitive processes and outcomes distinguishing these two types of interactions remain unexplored. Using functional near-infrared spectroscopy (fNIRS), a portable neuroimaging device, we investigate the neurobiological foundations of social connections initiated through conversations on both shallow and deep topics. By coupling neural-level activations with audiovisual recordings, we aim to explore the multimodal synergies across linguistic, behavioral, and neural dimensions during dyadic conversations. Utilizing hyperscanning and intersubject correlation analysis, we anticipate finding different activations in the pre-frontal cortex and temporal parietal junction (Lieberman, 2022). Through video analysis employing OpenFace and OpenPose—artificial intelligence algorithms that enable facial and gesture analysis—we aim to quantify the behavioral dynamics in dyadic interactions. By synthesizing methods commonly employed across various scholarly domains—including cognitive science, social psychology, and neuroscience—right from the idea inception stage, this study aims to explore possibilities for methodological integration. We focus on temporal correspondence (or time series) as the central element uniting these diverse approaches. Our goal is to make a meaningful contribution to the broader literature on multimodal synergy and to elucidate the complex interplay between mental and physical factors in the formation of social relationships.
Talk 3:	***Detection of freely moving thoughts using SVM and EEG signals*** Sairamya Nanjappan Jothiraj, University of Calgary, sairamya.nanjappanjo@ucalgary.ca, Julia Kam, University of Calgary, Julia.kam@ucalgary.ca, Caitlin Mills, University of Minnesota, cmills@umn.edu Freely moving thought is characterized as thoughts that shift from one topic to another without any prompts or overarching directions. As this phenomenon is often linked to creative thinking and positive mood, detecting when freely moving thought occurs can ultimately help improve our creative thought processes and mood. Despite its benefits, no studies to date have attempted to detect freely moving thought using electroencephalography (EEG) signals and machine learning approaches. This is the first study to our knowledge to examine the feasibility of using event-related potential (ERP) and spectral features of EEG signals in machine learning to detect freely moving thought. To address this aim, our classification models for detecting freely moving thought relied on previously collected EEG signals while performing a simple attention task. The statistical and entropy features of the P3 ERP and alpha spectral measures were entered as inputs to the support vector machine (SVM) for detecting freely moving thoughts. EEG features were first examined with both inter-subject and intra-subject strategies. The best combination of EEG features achieving higher classification performance in both strategies were then selected to combine with behavioral features to further enhance classification performance. Our best performing model has an MCC and AUC of 0.3105 and 0.6665 for inter-subject models and 0.2815 and 0.6407 for intra-subject models respectively. The above chance level performance in both strategies using EEG and behavioral features shows great promise for machine learning approaches to detect freely moving thought and highlights their potential for real-time prediction of freely moving thought. – withdrawn VISA issue
Talk 4:	***Evaluating Cognitive States and Trust in a Human-Machine Analytic Task*** Cara Widmer, Kairos Research, cara@kairosresearch.com, Amy Summerville, Kairos Research, amy@kairosresearch.com, Joshua Fiechter, Kairos Research,

	josh@kairosresearch.com Human analysts increasingly rely on automated teammates for making sense of vast troves of data. For these teammates to be useful, they should be both trusted by analysts to produce reliable and usable output, as well as reduce analyst's cognitive load. One feature of automated teammates that might facilitate both objectives is the degree of alignment between certain traits of their human counterparts. Participants completed an analytic task in which they examined supply chain data to make judgments about the criticality of companies to the network. They were assisted by an automated teammate that provided recommendations with varying degrees of detail. We examine the impact of recommendation type, as well as the degree to which recommendation type aligns with individual difference traits such as Need for Cognition and Locus of Control, on states of participant cognitive load and trust in the automated partner.
Talk 5:	***Generating Mastery: Developing a Closed Loop System to Support Mastery Learning*** Stephen Hutt, University of Denver, stephen.hutt@du.edu, Grayson Hieb, University of Denver, Grayson.Hieb@du.edu Mastery learning emphasizes a deep understanding of subjects, however, a common challenge in developing mastery learning materials is generating enough content that students are able to achieve mastery. This paper considers how advances in Generative AI may be leveraged to support mastery learning and learners more broadly. We detail the development and implementation of a proof-of-concept tool that leverages Generative AI to create accessible content generation and regeneration. Drawing upon open-access textbooks and online lecture transcripts as input, our system synthesizes summaries, crafts high-quality multiple-choice questions, and incorporates a re-explanation module tailored for students. For example, when reading lecture notes, students may interactively request that a section be rephrased, perhaps in different terminology. This will then be done in the context of the entire input (textbook, transcript, etc.). Quiz questions are evaluated against a rubric, employing both AI and human analysis, and achieved strong agreement between these assessments. This multifaceted approach allows students to engage with the learning material interactively, receive meaningful feedback, and request rephrasing or re-explanation of complex passages. Additionally, our system incorporates an auto-analysis mechanism to identify content themes and employs Generative AI to construct nuanced models of student knowledge. This enables the targeted delivery of personalized content and detailed information for instructors. Notably, the system has been designed with inclusivity in mind, catering to individuals with learning differences and non-native English speakers. We discuss the applications of this approach to scale up mastery learning techniques and support long-term encoding of knowledge.
Talk 6:	***Large Language Models and Human discourse processing*** Eyal Sagi, University of St. Francis, esagi@stfrancis.edu Recent advances in generative language models, such as ChatGPT have demonstrated an uncanny ability to produce texts that appear to be comparable to those produced by humans. Several key empirical results related to human processing of language, such as analogical reasoning, have been replicated using these models. Nevertheless, there are some important differences between the language generated by these models and language produced by humans. In this paper, I examine how an LLM performs on a sentence completion task reported on by Rhode, Levy, and Kehler (2011). This task examines how discourse relations, with a focus on an explanation relations, interact with syntactic ambiguities. While the completions offered by the LLM were all coherent, they differed from those provided by humans in key ways. Importantly, LLMs exhibited

	different biases for resolving syntactic ambiguity than humans. When completing a sentence such as 'Beth babysits the children of the jazz musician who…', humans tend to complete the sentence with reference to the jazz musician: '… lives in La Jolla'. In contrast, ChatGPT completed this sentence with reference to the children: '… have an innate sense of rhythm and music appreciation.'. This pattern was consistent across the stimuli. Nevertheless, ChatGPT also consistently chose explanation-based completions for verbs involving implicit causality (e.g., 'despises'), similarly to how humans complete these fragments. Because LLMs replicate language produced by humans, these results can help shed light on which aspects of language use are directly encoded in language and which require additional reasoning faculties beyond language processing.
Talk 7:	***Novel Neural Network Models for Predicting Mental Health Outcomes in the U.S. Youth Population*** Avi Verma, Palo Alto High School, vermaavi2006@gmail.com, Kaustubh Supekar, Department of Psychiatry and Behavioral Sciences, Stanford Medicine, ksupekar@stanford.edu Anxiety and depression are two of the most pressing mental health issues, particularly among young adults. Given the success of neural networks for predictive modeling, we developed novel neural network models for classifying anxiety and depression using Substance Abuse and Mental Health Services Administration's Mental Health Client- Level Data (SAMHSA MH-CLD) on 382,174 young adults (15 to 24 years old). The SAMHSA MH-CLD included mental health and general background data collected in 2020 for individuals reporting to state-accredited hospital service centers across the United States. The neural networks were trained on 10% randomized k-folds of the dataset and tested on the remaining 90% for each fold. We found that neural network models predicted anxiety and depression with high accuracy (91.5% to 94.2% accuracy, 8.4% to 3.1% loss), outperforming conventional statistical models. Additionally, for all tested variable sets, our neural network model outperformed expectations for the average therapist consultation (46% to 50% accuracy), as reported previously. For the optimal neural network model with the highest accuracy, the variables most correlated with anxiety and depression were age, education, gender, race, employment, marital status, and stressor events, not accounting for redundant and minimally correlated variables. The effectiveness of our neural network model indicates that it can be implemented alongside therapists in clinical environments to improve psychiatric diagnosis among young adults.
Talk 8:	Distributional language models and different kinds of semantic relations Jingfeng Zhang, University of Illinois Urbana Champaign, jz44@illinois.edu, Jon A. Willits, University of Illinois Urbana Champaign, jwillits@illinois.edu Distributional semantic models predict many behaviors, including categorization, semantic priming, semantic judgments, and ERPs. There are currently many different semantic models without a clear understanding of how they vary and which tasks and kinds of relations they predict well. For example, some models better predict paradigmatic/taxonomic relations, and other models better predict syntagmatic/thematic relations. Part of the difficulty in studying this is that it is hard to separate what effects are due to the model's architecture versus due to the corpora on which the model is trained. In the current study, we addressed this issue by training models on a carefully constructed artificial language corpus designed to allow for precise control of taxonomic and thematic relations. We used this corpus to train four models: 1) simple recurrent

	networks, 2) long-short-term memory networks, 3) word2vec, and 4) GPT-2. We showed that the four models varied systematically in terms of how well they learned the two types of relations, with the models that emphasize predicting precise word sequences (i.e., SRNs, LSTMs, and GPT-2) performing better (in that order) on thematic relations, and the model that encodes word order less precisely (word2vec) performing better on taxonomic relations. This work help demonstrates the strengths and weakness of different distributional language models, and how careful use of artificial corpora can help understand these models. The work also demonstrates how AI and machine learning models can help contribute to understanding of semantic memory.

Session 6 (10:30-12:30 PM): Internet Based Methods

table of contents *quick schedule*

Session:	Toolkits
Time:	10:30-12:30PM (*time set by organizer)*
Room:	Continental 3
Chair:	Ulf-Dietrich Reips
Symposium:	***News from Internet-based methods: Validation, replication, misinformation sharing, test scoring, Fish gaming, and AI*** The proposed symposium with six papers will bring together authors and a discussant from four institutions in Europe and the United States. It will continue the SCiP tradition of presenting methods, tools, and results for and from Internet-based research, with some looking into AI. Presentations concern the topics of Internet-driven generative AI, the psychology of sharing misinformation on the Internet, best scoring response methods in web-based knowledge assessment, Internet-based experimental methodology and replication, and studying the use of common goods via web-based interactive games. We present new methodological insights, new tools and new evidence that may help to improve Internet-based research methods and make them more widely known. There has been a lack of web-based implementations of common goods games that allow researchers to customize environmental parameters (e.g., cost-reward structure, speed of the game) and collect behavioral data. In the first presentation, Yury Shevchenko and Ulf-Dietrich Reips (University of Konstanz, Germany) introduce the open-source paradigm "Fish Lake" and explain how this web-based open-source software offers new opportunities for researchers of collective behavior. Tom Buchanan, Rotem Perach, and Deborah Husbands (University of Westminster, UK) investigate the Psychology behind the sharing of misinformation on the Internet. In their presentation “Why do People Share Political Misinformation Online? Findings and Methodological Challenges” they present two studies, in which they develop a framework for measuring sharing motivation and investigate behavioral predictors in individual differences and motivation. The third presentation, “Exact vs. Conceptual Replication: Mental Accounting in Internet-Based Experiments” by Maria Rosa Miccoli and Ulf-Dietrich Reips (University of Konstanz, Germany) highlights in-built advantages and limitations of Internet-based experimental methodology for replication. They report results from one exact and two conceptual replications of mental accounting. Birk Diedenhofen, Arvid Hofmann, and Jochen Musch (all University of Düsseldorf, Germany) present “Web-Based Knowledge Assessment: A Comparison of Multiple Choice, Constructed-Response and a new Axiomatically Derived Scoring Rule for the Subset Selection Response Format” as the fourth paper. They develop and test a promising new alternative response format, Zapechelnyuk's Axiomatic Scoring (ZAXS), by experimentally inducing different levels of knowledge in a large online study.

	In the fifth presentation, “Validating Google Ngram as a Source for Psychological Research: A Comparative Study of Corpora for Word Trend Analysis” Noemi Huber, Raphael Buchmüller and Ulf-Dietrich Reips (all University of Konstanz, Germany) scrutinize Google Ngram that has evolved as a research tool recently. For validation of psycho-cultural trends they analyze two different corpora and replicate results from the six most influential empirical articles that rely on Google Ngram methodology. The final presentation by Ulf-Dietrich Reips reveals how most of AI’s popular success really depends on the Internet and describes many of the dependencies like large corpora on the input side and worldwide interconnectivity for dissemination and improvement. He also discusses how AI may impact Internet-based research. The symposium will conclude with a summary provided by discussant and former SCiP president Christopher R. Wolfe (Miami University, Oxford, Ohio).
Talk 1:	***Fish Lake: A Web-Based Paradigm for Investigating Collective Behavior*** Yury Shevchenko (yury.shevchenko@uni-konstanz.de), Ulf-Dietrich Reips (reips@uni-konstanz.de), Research Methods, Assessment, & iScience, Department of Psychology, University of Konstanz, Germany Social interactions that involve cooperation and competition for common goods have been a topic of interest among researchers for a long time (e.g., Brucks et al., 2007). One way that researchers have studied collective behavior is through interactive games with paradigms such as the Tragedy of the Commons. While many studies have been conducted using these games, there has been a lack of web-based implementations that allow researchers to customize environmental parameters (e.g., cost-reward structure, speed of the game) and collect behavioral data. To address this gap, we introduce the open-source paradigm "Fish Lake". This is an online multiplayer game that enables real-time interactions between players. Using this online software, researchers can set up their own game sessions with customizable parameters to study individual and collective behavior via observation or experiment. To demonstrate the utility of Fish Lake, we will present a study that investigates the impact of game structure on participants’ behavior. Functions, availability and future developments of Fish Lake software will be described. Reference Brucks, W. M., Reips, U.-D., & Ryf, B. (2007). Group norms, physical distance, and ecological efficiency in common pool resource management. *Social Influence, 2*(2), 112–135. https://doi.org/10.1080/15534510701193436
Talk 2:	***Why do People Share Political Misinformation Online? Findings and Methodological Challenges*** Tom Buchanan (T.Buchanan@westminster.ac.uk), Rotem Perach (R.Perach@Westminster.ac.uk), Deborah Husbands (D.Husbands1@westminster.ac.uk), University of Westminster, UK Political misinformation on social media is a significant cause for concern around the world. It spreads in large part by human behaviour: on encountering false material, some people may share it to their own follower networks. Explaining why, and designing effective interventions, requires an understanding of how people who share misinformation differ from those who do not. Perhaps even more important to understand are any differences between those who do it by accident, believing the material to be true, and those who do it deliberately, knowing it is false.

	Thus, researching individual-level predictors of sharing misinformation is important. However, it is methodologically difficult, given the challenges of linking measurement of individual characteristics to actual behaviour 'in the wild'. This paper outlines two studies where we tried do this. In Study 1, we identified individuals who had shared false material on Twitter, and directly approached them to ask why. We used this to develop a framework for measuring sharing motivation. In Study 2, participants completed multiple measures of individual differences and motivation, and were asked for permission to examine their Twitter feeds. We coded the material they retweeted for presence of false political information, and linked this to previously-measured individual-level predictors. Multiple methodological problems were encountered in the course of this research: low response rates; labour-intensive data acquisition and coding processes; and concerns about researcher well-being. We will reflect on ways in which computational approaches including AI could be used to address these issues.
Talk 3:	***Exact vs. Conceptual Replication: Mental Accounting in Internet-Based Experiments*** Maria Rosa Miccoli (maria-rosa.miccoli@uni-konstanz.de), Ulf-Dietrich Reips (reips@uni-konstanz.de), Research Methods, Assessment, & iScience, Department of Psychology, University of Konstanz, Germany Replication research in psychology would benefit from rigorous application of some methodologies and transparent processes characterizing Internet-based research. Since the replication crisis, questions have been raised about the internal validity of replication studies: often, the lack of complete reports about the procedure and materials used in laboratory studies hampers full insights into the cause-and-effect relationship established in previous studies. Conducting Internet-based research studies helps in countering this issue of internal validity, because materials are directly available and copyable on the Internet. Also, replication studies often need larger sample sizes than the original ones, due to the differences between intended and actual power in replication studies (Anderson & Maxwell, 2017), and Internet-based research offers easy access to large samples. We discuss the advantages of conducting replication studies with Internet-based methodologies separately for exact and conceptual Internet-based replication studies. “Exact” (Hudson, 2023) replication studies are performed to check for internal validity, while conceptual replication studies assess the external validity of the psychological effects. We present three Internet-based experiments, including one exact replication and two conceptual replications testing the mental accounting effect based on Tversky and Kahneman's classic paradigm (1984). The Internet-based version of the exact replication maintains the same experimental design as the original paradigm. The two conceptual replications are manipulated within- and between-subjects, with the aim of assessing the mental accounting effect when varying the experimental process components. After presenting the results, we discuss the methodologies of Internet-based research that best assess the replicability of psychological effects. References Anderson, S. F., & Maxwell, S. E. (2017). Addressing the “replication crisis”: Using original studies to design replication studies with appropriate statistical power. *Multivariate Behavioral Research, 52(3)*, 305-324. https://doi.org/10.1080/00273171.2017.1289361

	Hudson, R. (2023). Explicating exact versus conceptual replication. *Erkenntnis, 88,* 2493 -2514. https://doi.org/10.1007/s10670-021-00464-z Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist, 39(4)*, 341–350. https://doi.org/10.1037/0003-066X.39.4.341
Talk 4:	***Web-Based Knowledge Assessment: A Comparison of Multiple Choice, Constructed-Response and a new Axiomatically Derived Scoring Rule for the Subset Selection Response Format*** Birk Diedenhofen (birk.diedenhofen@uni-duesseldorf.de), Arvid Hofmann (arvid.hofmann@uni-duesseldorf.de), Jochen Musch (jochen.musch@uni-duesseldorf.de), Department of Psychology, University of Duesseldorf, Germany We evaluated classical and novel methods for assessing knowledge online. Multiple-choice (MC) tests have been found to permit a reliable and valid assessment of knowledge, but can be criticized for encouraging guessing and failing to capture partial knowledge. The open-ended constructed-response (CR) format requires active recall rather than mere recognition, but scoring is cumbersome and objectivity cannot be taken for granted. A new test format that aims to overcome the drawbacks of existing methods is Zapechelnyuk's Axiomatic Scoring (ZAXS). It combines a subset selection response format that allows test takers to choose k out of the n possible options, and a scoring rule that grants partial credit by awarding 1 / k points if the selected options include the correct answer. To compare the three competing test formats, we established a strong validation criterion by experimentally inducing different levels of knowledge in a large online study. We found ZAXS scores to be more reliable than MC scores, but less reliable than CR scores. In terms of validity, ZAXS scores reflected the respondent's known knowledge better than both CR and MC scores. We conclude that ZAXS is a promising new alternative response format and scoring method for web-based knowledge assessment.
Talk 5:	***Validating Google Ngram as a Source for Psychological Research: A Comparative Study of Corpora for Word Trend Analysis*** Noemi Huber, Raphael Buchmüller, Research, Methods, Assessment, & iScience, Department of Psychology, University of Konstanz, Germany, Ulf-Dietrich Reips, Data Analysis & Visualization, Department of Computer and Information Science, University of Konstanz, Germany The Google Books Ngram Viewer (GNV, https://books.google.com/ngrams/) is a large linguistic online corpus created by Google, consisting of over 500 billion words from the years 1500 to 2019 in eight different languages. The GNV is used in psychological and cultural trends research to analyze the occurrences of selected words over time (Michel et al., 2011). However, researchers question the validity of the GNV to represent actual cultural phenomena. No conclusive research on the validity of the GNV method has been conducted so far. To begin with the validation of the GNV method, we address the above concerns by replicating the research in six publications that originally make use of the GNV for trend analysis. We apply the publications' methods and procedures with similar corpora, namely the Corpus of Historical American English (COHA, https://www.english-corpora.org/coha/) and the TIME Magazine corpus (https://www.english-corpora.org/time/), and analyze the replicability of the results. We investigate the individual and average correlations between word occurrences and time and conduct a visual assessment to identify temporary phenomena such as turning points. Following the guidelines by Younes and Reips (2019) to increase the quality of GNV-based results, we implement synonym-use, inflection analyses and standardization procedures to improve the accuracy of the results.

	Results between the three corpora turn out to be similar, with some predictable differences related to the type of corpus. We discuss implications for the use of the GNV method and conclude that using the American Google Ngram viewer is valid for psychological research and that previous findings from GNV-based research are supported. References Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., the Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644 Younes, N., & Reips, U.-D. (2019). Guideline for improving the reliability of Google Ngram studies: Evidence from religious terms. *PLOS ONE*, *14*(3), e0213554. https://doi.org/10.1371/journal.pone.0213554
Talk 6:	***Without the Internet Artificial Intelligence was Disconnected Stupidity*** Ulf-Dietrich Reips (reips@uni-konstanz.de), Research Methods, Assessment, & iScience, Department of Psychology, University of Konstanz, Germany Generative artificial intelligence is a branch of artificial intelligence (AI) that can create new content, such as images, text, music, or videos, based on existing data. As an example, the present abstract was generated with assistance by generative AI. Decades after its principal invention, generative AI suddenly (re)appears and captures public attention and imagination. However, the success of generative AI was only possible because of Internet technologies that enabled the *collection* of massive amounts of data, as well as the *distribution and access* of AI tools and generated content. The Internet allows users to access generative AI services like any *web services* (e.g. Reips & Lengler, 2005) through web or mobile applications, without having to install or maintain any software or hardware. Another important Internet technology that contributed to the success of generative AI is *social media*, which provide a rich source of data for generative AI models to learn from, as well as platforms for sharing and consuming the generated content. Social media also enable feedback and interaction between users and generative AI systems, which can improve the quality and diversity of the content, but may raise ethical and legal questions. The Internet further fosters the emergence and growth of open-source platforms, libraries and communities, such as TensorFlow, PyTorch, Hugging Face, and OpenAI, that provide access to pre-trained generative AI models, datasets, tools, and frameworks, as well as foster collaboration and innovation among researchers and developers (Chui et al, 2023; Gartner, 2023). Powered by the Internet the increasing demand and adoption of generative AI applications then in turn creates quick growth. I will discuss implications for AI from its reliance on the Internet, and how AI may influence Internet-based research. References Chui, M., Roberts, R., Rodchenko, T., & Singla, A. (2023). *What every CEO should know about generative AI*. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/what-every-ceo-should-know-about-generative-ai Gartner (2023). *Gartner experts answer the top generative AI questions for your enterprise*. https://www.gartner.com/en/topics/generative-ai Reips, U.-D., & Lengler, R. (2005). The Web Experiment List: A Web service for the

	recruitment of participants and archiving of Internet-based experiments. *Behavior Research Methods, 37*, 287-292. https://doi.org/10.3758/bf03192696
Discussion	Christopher R. Wolfe, wolfecr@muohio.edu, Miami University, Oxford, Ohio, USA

Session 7 (1:30-2:45 PM): Language

table of contents *quick schedule*

Session:	Language
Time:	1:30-2:45PM (*12 minutes + 3 minutes Q&A)*
Room:	Continental 1-2
Chair:	Laura Allen
Talk 1:	***Comparing humor norms of English words across speakers of North American, British, and Singapore English*** Cynthia S. Q. Siew, National University of Singapore, cynthia@nus.edu.sg Large-scale collection of lexical-semantic norms for words in a given language has been instrumental in the progress of psycholinguistic research. However, such norms tend to be collected from speakers of the dominant variant or dialect. This research aims to determine if there may be differences across speakers of various dialects of English in the humor of individual words. Engelthaler and Hills (2018) observed that their humor ratings were most strongly correlated with inverse word frequency: Less frequent words tended to be rated as more humorous. We hypothesized that words that are less frequently occurring in a given English dialect should be perceived as more humorous by speakers of the same dialect. We selected words of relatively higher and lower frequencies across various corpora of North American, British, or Singapore English, and presented these words to participants who were native English speakers of North American, British, or Singapore English. Study 1 compared humor ratings of North Americans and Singaporeans; Study 2 compared humor ratings of North Americans and the British. Humor ratings were generally more strongly (and inversely) associated with the word's frequency in the corpora that aligned with the rater's English dialect. Our results provide support for the idea that people are sensitive to the statistics of their specific language environment, and importantly suggest that creators of lexical-semantic databases should consider how the cultural, historical, or sociopolitical context of raters could influence the nature of their ratings.
Talk 2:	***Aphantasia Explored: Insights from Online Communities*** Püren Öncel, University of Minnesota, oncel001@umn.edu, Laura K. Allen, University of Minnesota, lallen@umn.edu Experiences of visual imagery vary greatly among individuals, with some people regularly conjuring vivid mental images while others report experiencing only darkness (Keogh & Pearson, 2018). Aphantasia, affecting 2-3% of the population, denotes the inability to form mental images. Despite its prevalence, limited research has examined the psychological implications of this condition. This study focuses on a subreddit devoted to Aphantasia (r/Aphantasia), where personal experiences are shared anonymously. We employed topic modeling to uncover the perspectives of individuals who experience this condition. Analysis revealed 35 topics from users' posts, reflecting the multifaceted experiences of aphantasic individuals. Consistent with prior literature (e.g., Keogh et al., 2021), the themes relate to aphantasia's effects on creativity, mental health, memory, auditory imagery, spatial ability, and sensation. Notably, our findings also unveiled novel topics

	not commonly discussed in the literature – namely, individuals discussed a wide range of experiences related to their processing of narratives and linguistic content more broadly. The findings from this study emphasize crucial factors for future empirical research in this field. In conclusion, our study's topic modeling approach uncovers nuanced topics within the Aphantasia subreddit, providing valuable insights into the varied experiences of Aphantasic individuals. By delving into unexplored dimensions adopting ecologically valid methods, we expand our understanding of Aphantasia and its psychological impact, particularly revealing critical insights into narrative experiences and language. This research contributes to the broader discourse on the diversity of cognitive experiences by highlighting examining factors that may influence individuals' experiences in the world.
Talk 3:	***Embodying verbs with a database of 3-D motion norms*** John Hollander, University of Memphis, jmhllndr@memphis.edu, Andrew Olney, University of Memphis, aolney@memphis.edu Psycholinguistic studies are increasingly considering the role of embodied information in language processing. Reflecting this trend, researchers are recently building lexical databases that include perceptual or sensorimotor information, such as sensory modality strength, manipulability, and spatial localization, with an emphasis in nouns over other parts of speech. In this talk, we report our efforts to build an embodiment-focused lexical database tailored to verbs, especially (but not limited to) those that imply physical motion. We present analyses of the directionality and strength of motion implied by 320 verbs in three dimensional space. Data were obtained with surveys of human participants. We sought to assess the convergent and divergent validity of our database by comparing its semantic space to spaces derived from other psycholinguistic norms and distributional semantic techniques (e.g., Word2vec). To do so, we obtained all pairwise cosine similarities between the word vectors in our data to create rank ordered lists of semantic neighbors. We then compared these lists to the semantic neighborhoods derived from other psycholinguistic databases to generate distributions of Spearman's rho correlations, which describe the similarity of the overall semantic spaces. Results suggest that this norming method is supported by convergent validity while capturing unique psycholinguistic information. Further, lexical databases derived from more perceptually-oriented norming tasks yielded more similar semantic space to our data than distributional semantic models. These norms may contribute to stimulus selection and refinement, generating computational representations of verbs, and novel research designs in future studies.
Talk 4:	***Using Neural Language and Vision Models to Demonstrate Task Specific Activation of Different Semantic Information*** Andrew Z. Flores, University of Illinois Urbana Champaign, azf2@illinois.edu, Jon A. Willits, University of Illinois Urbana Champaign, jwillits@illinois.edu One recent question regarding semantic memory is the extent to which different semantic tasks activate different semantic information. In the current study, we conducted three experiments with different tasks, and predicted the behavioral responses using different measures and models of semantic relatedness that index different kinds of semantic information. All three experiments used the same item pairs belonging to four related conditions: 1) most-related (tiger-lion), 2) same category (tiger-moose), 3) same artifact/natural kind category (tiger-butterfly), 4) different artifact/natural kind category (tiger-shoe). Experiment 1 was an eye tracking experiment where participants were shown two pictures and asked to look at one of the pictures.

	Experiment 2 participants were asked to click on the named picture. Experiment 3 asked the participants to make a semantic similarity judgement between the two items. All three experiments showed significant differences across the four conditions in expected directions (more interference or higher similarity rating as a function of similarity category). We then modeled the individual item pair data from all three experiments using 1) low-level measures of visual similarity, 2) similarity in sensory-motor norms, 3) similarity in convolutional neural networks, and 4) similarity in distributional language models. The different predictors had markedly different effects across the three experiments. The experiments demonstrate how different kinds of semantic information (such as low level visual and motor knowledge, categorization knowledge, and linguistic similarity) may contribute differently to different semantic tasks. The experiments also demonstrate how AI and machine learning models can help contribute to understanding of semantic memory.
Talk 5: (withdrawn)	***The Commonalities Between Deja Vu, Involuntary Autobiographical Memories, and Unexpected Thoughts: A Multidisciplinary Approach to Investigating Phenomenology Across the Lifespan*** Cati Poulos, University of Minnesota, poulo032@umn.edu, Videep Venkatesha, Colorado State University, videep@colostate.edu, Caitlin Mills, University Of Minnesota, cmills@umn.edu, Nathaniel Blanchard, Colorado State University, Nathaniel.Blanchard@colostate.edu, Anne Cleary, Colorado State University, Anne.Cleary@colostate.edu Involuntary thoughts are commonplace in day-to-day life, and their forms are quite varied, as can be seen in the examples of déjà vu, involuntary autobiographical memories (IAMs), and unexpected thoughts. Prior research suggests that these phenomena have common characteristics, such as having some form of peak in young adulthood relative to other points along the lifespan, but how they may be similar versus distinct from one another is not well-understood. In the present study, we examined the phenomenological appraisal patterns of deja vu, IAM, and unexpected thoughts in a sample of older and younger adults to address this gap. We took a multidisciplinary approach by combining null hypothesis testing with natural language processing techniques. Our results suggest that there are distinct features in the descriptions of IAM, deja vu, and unexpected thoughts, and separability of language used by older adults and younger adults. Older adults' phenomenological experiences of involuntary thoughts also differed significantly from younger adults' experiences across various dimensions. These results contribute to our developing understanding of the umbrella of thought spontaneity by investigating how involuntary thoughts change in experience as we age and how these changes vary as a function of the type of thought.

Session 8 (1:15-2:45 PM): Thinking

table of contents *quick schedule*

Session:	Thinking
Time:	1:15-2:45PM (*12 minutes + 3 minutes Q&A)*
Room:	Continental 3
Chair:	Erin M. Buchanan
Talk 1:	***An Investigation into How Native Spanish Speakers Who Learned English as a Second Language Understand the Gist of Complex Medical Texts in English*** Josselyn E. Marroquín, Miami University, marroqje@miamioh.edu, Christopher R. Wolfe, Miami University, wolfecr@miamioh.edu Little research has focused on native Spanish speakers who speak English as a Second Language (ESL) regarding how to write medical information to promote gist comprehension. In this study, 181 ESL Hispanic/Latine/a/o native Spanish speakers were recruited from across the United States. An “authentic” article in English about universal flu vaccines was taken from the web, analyzed with Coh-Metrix, and further analyzed using Gist Inference Scores (GIS), a measure of how likely people are to understand a text’s bottom-line meaning. The article was revised to obtain a higher GIS. Participants were randomly assigned to the original low GIS article, the improved high GIS version of the article, or a control article. Then, participants were asked questions about the flu vaccines using 7-point Likert scale questions to assess gist comprehension, and multiple-choice questions to assess verbatim knowledge conveyed in both the low GIS and high GIS articles about universal flu vaccines. In order to test differences in a person’s health literacy, participants also filled out a health literacy questionnaire. Results found that there were no significant differences between the groups for gist comprehension and verbatim knowledge multiple-choice questions. Groups did not differ in health literacy and did not predict other outcomes. We discuss limitations in recruiting Hispanic/Latine/a/o ESL participants, and strategies to improve data quality in future randomized, controlled web-based experiments with this underrepresented population.
Talk 2:	***Identifying Déjà Vu: An Automatic Approach Using Eye Gaze Features*** Iliana Castillon, Colorado State University, ilianaca@rams.colostate.edu, Videep Venkatesha, Colorado State University, videep@rams.colostate.edu, Anne Cleary, Colorado State University, Anne.Cleary@colostate.edu, Nathaniel Blanchard, Colorado State University, Nathaniel.Blanchard@colostate.edu Long term, déjà vu is linked to cognitive processes like familiarity detection, curiosity, and internal memory search. In this work, we investigate the feasibility of automatically identifying when someone is experiencing déjà vu from eye-gaze features. We collected a dataset of users both experiencing and not experiencing déjà vu, extracted eye-gaze features from each group, and assessed the relationship between these eye-gaze features and the occurrences of déjà vu. We establish a clear link between eye-gaze feature and the déjà vu phenomenon, and that déjà vu can be automatically detected

	using eye-gaze features with a Cohen's Kappa of 0.22. Given past work suggesting that déjà vu may direct attention inward toward memory search, this may provide a potential method of determining when attention has shifted from being externally oriented to being internally oriented.
Talk 3:	***Knowledge Structures in Psychology: Are All Sub-Areas the Same?*** Ruth S. Day, Duke University, ruthday@duke.edu Psychology is a diverse discipline. It has many sub-areas that vary in their reliance on natural science vs. social science approaches, as well as types of phenomena and theories. Do these differences affect the knowledge structures of people? Do initial knowledge structures change with more experience in the discipline? We examined knowledge structures across several areas of psychology: Biological, Cognitive, Developmental, Social, Clinical, Methods, and Omnibus (the entire discipline). Participants varied in their amount of prior knowledge and experience in the discipline – introductory psychology students, their teaching fellows, and their instructor. For each area, participants saw a list of key concepts (selected by the instructor from the course) and sorted them into piles based on perceived similarity. The results were displayed in tree diagrams and scatterplots, showing the relatedness of items. We used new computational tools to determine the amount of structure in each area as well as the type of structure. The overall amount of structure was surprisingly low for the students, even though their performance in the course was very good. Therefore students had acquired considerable knowledge but had not developed systematic structures for that knowledge. There were other surprises. For example, an area of considerable a priori student interest yielded very little structure at all while another yielded strong structure. These results suggest that studying the knowledge structures of people can also provide insights about the underlying structures of the discipline itself. Also, differences between the students and their teaching staff suggest implications for teaching and learning.
Talk 4:	***Thinking Grid: A Tool to Measure Thought Dynamics*** Vishal Kuvar, University of Minnesota, kuvar001@umn.edu, Samuel Murray, Providence College, smurray7@providence.edu, Mya Urena, University of Minnesota, urena014@umn.edu, Caitlin Mills, University of Minnesota, cmills@umn.edu, Zac Irving, University of Virginia, zci7c@virginia.edu The study of consciousness has typically focused on the content of thought, neglecting its dynamics, or how thinking unfolds over time. This is partly due to methodological limits on accessing the stream of consciousness. To fill this gap, we validated a measure of thought dynamics, called the Thinking Grid(TG), based on Christoff et al.'s (2016) view that the dynamics of thought depend on the degree of automatic and deliberate constraints. Accordingly, the TG consists of two axes: (a) deliberate constraints, capturing the influence of cognitive control, and (b) automatic constraints, capturing the influence of affectively- or perceptually-salient stimuli. Participants read six vignettes in which a proxy is experiencing freely-moving (FMT; moving from topic to topic), directed (DT; deliberately focusing) or sticky thoughts (ST; repeatedly coming back to some topic). Each thought type was crossed with task constraints, where the proxy had a focal task or not. Following each vignette, participants reported where the proxy's thoughts fell on the TG. Responses on this space were converted into clusters, where each cluster was assigned a thought type depending on where it fell on the grid. Results indicated that participants' responses on the TG were significantly more likely to fall in the FMT cluster when the proxy's thoughts moved from one topic to another.

	Similar significant effects were also seen on ST and DT conditions, and there was no effect of task constraints. These results indicate that participants can intuitively classify experiences on the TG, providing a novel method to measure consciousness.
Talk 5:	***Computational approaches to studying streams-of-consciousness*** Constance Bainbridge, UCLA, cbainbridge@ucla.edu, Rick Dale, UCLA, rdale@ucla.edu Just as language facilitates communication with social others, it also permits internally generated streams-of-consciousness. At any time, one's thoughts may be influenced by those that came before, and these trajectories can then shape how we think about different topics, such as considerations of the future. The language contained in such streams-of-consciousness may also be revealing of mental health or wellness attributes, providing clear motivation to better understand our spontaneous thoughts and how they are shaped. Here, we offer a methodological survey of computational techniques that can be used to understand what influences our thought trajectories, and what they can reveal about us. This includes classic content-based methods such as Linguistic Inquiry Word Count (LIWC) collapsed into conceptual components. There are also methods that quantify timing of thoughts, such as analysis of the temporal intervals between keystrokes or spoken utterances. To illustrate one application, we will share work we have completed exploring how temporal anchoring in the past or present distinguished the topical spaces explored about the future when thinking about life surrounding the COVID-19 pandemic. We conclude with future possible applications, such as the use of Large Language Models (LLMs) to measure, generate, or alter typed streams-of-consciousness.

Poster Session (12:30-1:30 PM)

table of contents *quick schedule*

Session:	Poster Session
Time:	12:30-1:30PM (*please set up before lunch!)*
Room:	Grand Ballroom
Poster 1:	Improving Text Simplification in a Real-World Context: Leveraging the Automatic Readability Tool for English (ARTE) Kathryn S. McCarthy, Georgia State University, kmccarthy12@gsu.edu, Joon Suh Choi, Georgia State University, jchoi92@gsu.edu, Scott A. Crossley, Vanderbilt University, scott.crossley@vanderbilt.edu The Automatic Readability Tool for English (ARTE) provides multiple traditional and theory-informed readability scores for texts. ARTE is available as a website and as a downloadable, desktop tool. The website includes a transformer-based readability formula and users can input a text and receive recommendations of texts from the CommonLit Ease of Readability corpus that are similarly difficult based on different readability metrics. In this poster, we introduce ARTE and then demonstrate its utility in a civics literacy context. Roll-off, or skipping items on a ballot, is a major threat to the democratic process. Prior correlation work suggests that the linguistic complexity of ballot measures (local issues voted on directly by the people) is a strong predictor of roll-off. In the study, we used ARTE to guide three types of text simplification (surface, Plain Language, discourse-theory informed). We used an experimental design to test if the three types of simplification had differing effects on participants' (n = 101) perceived comprehension (i.e., metacomprehension) and whether they voted or chose to roll off. The results showed that all three simplification types had improved metacomprehension ratings and reduced roll-off relative to the original ballot measures. Importantly, the discourse-informed simplification had stronger effects than simplification that relied on revisions that were guided by traditional readability metrics. This study suggests that ARTE can serve as a promising step toward automated or tech-augmented approaches that can support more effective text simplification. Such advances could help make complex real-world texts more accessible for more people.
Poster 2:	***Usability Testing of Power Analysis Software Prototypes*** Jeongyun Choi, Purdue University, choi660@purdue.edu, Ya-Hsin Hung, Purdue University, hung17@purdue.edu, Robert W. Proctor, Purdue University, rproctor@purdue.edu, Erin P. Hennes, University of Missouri, ephyr8@missouri.edu, Sean P. Lane, University of Missouri, lanesp@missouri.edu Power analysis is a key strategy for a priori sample size determination, and scholars are increasingly developing software tools to assist in these calculations. Power analyses consist of (1) determining a planned statistical model, (2) estimating each model parameter, (3) calculating and (4) interpreting power estimation results via simulation or closed-form solution. In the current research, we conducted usability tests contrasting alternative prototypes for three of these steps. For model selection, the prototypes depicted model-based (e.g., Webpower), family-based (e.g., G*Power), and

	procedure-based user interfaces (UI; e.g., PASS). For parameter estimation, the prototypes were in a one-at-a-time, pop-up window (available in G*Power), expandable panel (e.g., JMP), and tab-page (e.g., PASS). For results, the prototypes provided basic browsing, exported view (available in SPSS), comparison (available in nQuery), and hybrid views (e.g., PIFACE). Participants inspected and used the prototypes to accomplish task goals while being video recorded, completed the System Usability Scale (SUS), and were interviewed. For model selection, the procedure-based UI was deemed most usable. For parameter estimation, the expandable panel style was deemed the most usable. For results interpretation, the basic dropdown browsing view was deemed the most usable. However, explicit preferences did not always match SUS rankings. Content analysis categorized the most crucial issues as user knowledge, structure/organization of the UI, and intuitiveness of the UI for model selection, parameter estimation, and results view, respectively. This study will contribute to user-friendly power analysis tool development by empirically validating both what users prefer and what factors contribute to correct tool usage.
Poster 3:	***Snap, crackle, pop: Motion and motion derivatives in mousetracking data*** Stephanie Huette, University of Memphis, shuette@memphis.edu, David Heath, University of Memphis, daheath@memphis.edu Many mousetracking studies utilize spatial metrics and velocity for dependent variables, but there exist several other informative markers of movement within movement data. Higher order derivatives of motion such as jerk are indeed experienced by us and may provide a useful measurement of underlying cognitive processes. In this work, participants completed a difficult maze using a computer mouse. One quadrant of the maze had several plausible routes and led to back-tracking, while another part of the maze was straightforward to find the path through after getting through the difficult area. The difficult area and easier areas were compared in terms of stop time, velocity, and all higher order derivatives of motion (acceleration, jerk, snap, crackle, and pop). While no significant differences were found in velocity or stop times, higher order derivatives all showed robust significant differences, underscoring a need to further explore and operationalize these measurements within various experimental frameworks. Discussion of these motion differences along with experiential effects of higher order motion derivatives and possible connections to cognitive processes will be discussed.
Poster 4:	***SCIP: Combining group communication and interpersonal positioning to identify emergent roles in scaled digital environments*** Nia Nixon, University of California, Irvine, Dowelln@uci.edu, Sasha Poquet, Technical university of Munich, sspoqut@gmail.com We propose a novel approach to the assessment of the emergent socio-cognitive roles learners adopted during peer interactions. The approach posits that different dimensions of peer interaction emerging from temporal-semantic discourse information and the structure of interactions can be used to diagnostically reveal the emergent roles of learners during peer interactions. As such, the combination of two established methodologies, Group Communication Analysis (GCA) that centers on temporal semantic properties of online discourse with Social Network Analysis (SNA) that reflects structural interpersonal patterns of online interactions are used to gain a deeper understanding of the emergent, socio-cognitive roles learners adopt during peer interactions at scale. The proposed approach is named socio-cognitive group communication and interpersonal position (SCIP) analysis and is defined as a combination of these two distinct and complementary analytic techniques. The proposed SCIP approach is examined on data produced during peer interaction in a massive open online course (MOOC) delivered via Coursera. Using SCIP analysis, learner activity is described through five roles: Lurkers,

	Followers, Socially Detached, Influential Actors and Hyper Posters. We conclude the paper with a detailed discussion of the theoretical, methodological, and practical implications for peer interaction research. The scalability of the methodology opens the door for future research efforts directed towards understanding and improving peer-interactions at scale.
Poster 5:	***Collaboration Over Time as Iterative Bayesian Inference within a Dynamical Systems Model*** Grace Qiyuan Miao, University of California, Los Angeles, q.miao@ucla.edu, Andrew Jun Lee, University of California, Los Angeles, andrewlee0@g.ucla.edu, Hongjing Lu, University of California, Los Angeles, hongjing@ucla.edu, Rick Dale, University of California, Los Angeles, rdale@ucla.edu, Alexia Galati, University of Carolina, Charlotte, agalati@uncc.edu People collaborate in dyads to perform joint tasks every day, and computational scholars have developed different models to capture human interactional dynamics. We recently proposed a model in the dynamical systems tradition wherein the change in behavior of dyads over time is governed by a latent variable that describes the kind of interaction in the dyad (Miao, Dale, & Galati, 2023). Crucially, different values of this variable, called the "context matrix," qualitatively correspond to different interaction states that range from behavioral synchrony to complete randomness. However, two limitations confront the model. First, all values of the context matrix are discrete, so that there are no intermediate degrees of interaction states. Second, no mechanism allows for change in the context matrix for the same dyad in the same task. Here, instead of modeling just the effect of dyadic interaction in subsequent behavior as in the original model, we propose that change in behavior iteratively updates a distribution of possible states of dyadic interaction according to Bayes' theorem. At every time point, the average context matrix of this updated posterior distribution is selected to induce the next behavioral prediction. We select the average matrix, as opposed to the most likely one, as it allows us to set continuous values in the matrix and to weigh probabilistic uncertainty. Our results explore the possibility of integrating two major computational traditions – dynamical systems and Bayesian inference – to address cognitive questions related to human collaboration.
Poster 6:	***A watched clock never ticks? Mind wandering and boredom in the presence and absence of a clock.*** Jorge Horan-Barnes, Tilburg University, j.o.horan-barnes@tilburguniversity.edu, Mariana Dias da Silva, Tilburg University, M.R.DiasDaSilva@tilburguniversity.edu, Myrthe Faber, Tilburg University, m.faber@tilburguniversity.edu When feeling bored, it is common to focus on time passing, with minutes feeling like hours (Danckert & Allman, 2005). Previous work on mind wandering has found that indeed, people report thoughts about the passage of time (e.g., thinking about "how long I have been reading for" or "how much longer the reading is and what time it is"; examples from Faber & D'Mello, 2018) when zoning out. In this study, we asked whether having access to a clock influences mind wandering and feelings of boredom. On the one hand, clockwatching can make people more aware of the passage of time, and therefore mind wandering and feelings of boredom might increase. On the other hand, a recent study found that the provision of live bus times reduced passengers' perceived waiting time by 13% (Lu et al., 2018), suggesting that people might in fact focus less on the passage of time when a clock is present. To study the effect of clock presence on boredom and mind wandering, participants performed web-based reading of a boring and an exciting narrative text (paradigm similar to Danckert et al., 2018), with a clock either present or absent on the computer screen. We recorded their eye-movements, reading comprehension, ratings of boredom, and mind wandering probes. In this presentation, we

	will report on the findings of this study. The findings are important for understanding whether the presence of a clock can help reduce mind wandering, and can inform decisions on the visual presentation of computerized tasks.
Poster 7:	***Improved knowledge when using comics versus texts in education*** Marianna Pagkratidou, University of Minnesota, mpagkrat@umn.edu, Neil Cohn, Tilburg University, neilcohn@emaki.net, Phivos Phylactou, University of Western Ontario, pphylact@uwo.ca, Marietta Papadatou-Pastou, National and Kapodistrian University of Athens, marietta.papadatou@gmail.com, Gavin Duffy, TU Dublin, gavin.duffy@tudublin.ie The past decades have seen a growing use of comics education material in STEM and non-STEM related fields. However, there are inconsistent reports regarding the effectiveness of learning when using comics—the use of visual language and writing in sequential images that convey education material—compared to texts. In this study, we conducted a meta-analysis to quantify the overall effect of comics vs texts that have been used in empirical studies that targeted learning for STEM and non-STEM fields. Results showed an overall moderate effect in favour of comics compared to texts in STEM and non-STEM learning, indicating increased knowledge due to the use of comics. Findings which not only contribute to the Visual Language Theory and spatial cognition, but also shed light on the conditions in which comics can foster learning.
Poster 8:	***Decoding Narrative Discourse: A Computational Exploration of Media Modality Effects*** Shu Hu Georgia State University, shu13@gsu.edu, Puren Oncel, University of Minnesota, Twin Cities, puren.oncel@gmail.com, Heather Ness-Maddox, Middle Georgia State University, heather.nessmaddox@mga.edu, Laura Allen, University of Minnesota, Twin Cities, laura.allen22@gmail.com, Joseph P. Magliano, Georgia State University, jmagliano@gsu.edu We experience narratives across a variety of media, such as texts, comics, and film. While there are common processes that support how narratives are understood across media, there are differences in the affordances of media that could affect how narratives are understood. More research is needed to explore how the affordances of different media affects the way people understand events. In this study, we explored this issue in the context of a think-aloud task, which provides a basis for how media affects the way people convey their understanding in language. Based on claims in the literature about differences in the affordances of media on situation models, it was hypothesized that there would be differences across media in the extent situational information was conveyed. Specifically, it was hypothesized that text afforded the representation of characters' internal states (emotion and cognitive) more so than visual narratives, whereas visual narratives afforded spatial information more than texts. Participants produced typed think-aloud responses while reading text or picture versions of the same stories. Natural language processing tools were used to assess the presence of situational indices in the think-aloud responses. Specifically, LIWC and general inquirer were used to assess the extent that participants used words reflecting character internal states and spatial information. Participants produce more words about the internal states of characters in the text condition than in the visual condition. In contrast, participants produce more place words in the visual condition than in the text condition. Implications for theories of comprehension will be discussed.
Poster 9:	***An extensive dataset evaluating breadth and desirability for 1,214 adjectives*** Lin Lin, University of California, Los Angeles, llin001@g.ucla.edu, Rick Dale, University of California, Los Angeles, rdale@ucla.edu, Steve Stroessner, University of California, Los Angeles, ss233@g.ucla.edu

	Language plays a central role in a variety of fields, reflecting its fundamental importance in human communication, social interaction and cognition. People exhibit subtle biases through language, often without conscious awareness or intent. Indeed, the breadth of language used to characterize an event or an object can influence how it is interpreted and has psychological and behavioral implications. According to the Linguistic Category Model (LCM), the same event or behavior could be described with different levels of breadth. As linguistic breadth increases, more information is conveyed, and it is more challenging to verify the veridicality of information. Evidence supportive of the LCM has amassed over decades, yet focused almost exclusively on differences in verb usage. A neglected question involves whether similar effects arise in using adjectives to characterize actions and events. Adjectives are central means of communicating detail (e.g., big, old, red) and, when describing social groups, constitute the semantic content of stereotypes (e.g., lazy, rude, aggressive). This work reports a new database which contains ratings for 1,214 English adjectives on the dimensions of breadth and desirability. Approximately 100 individual ratings were collected for each word, providing a stable current estimate of these two dimensions for each. For example, the words “good” and “normal” have the highest score on breadth, while “unpunctual” and “untalkative” have the lowest score. This database will help to refine analysis of semantic breadth and desirability in communications, aid studies on text analysis, and semantic representation in brain and behavior pertaining to communication.

Presidential Symposium (3:00-4:00PM)

table of contents *quick schedule*

Session:	Presidential Symposium
Time:	3:00-4:00PM *(15 minutes + 5 minutes Q&A)*
Room:	Continental 1-2
Chair:	Caitlin Mills
Talk 1:	Julia Kam
Talk 2:	Myrthe Faber
Talk 3:	***Language networks as a framework for exploring the mental lexicon*** Cynthia Siew

Keynote Speaker (4:00-5:00 PM)

table of contents *quick schedule*

Session:	Keynote Speaker
Time:	4:00-5:00PM (*45 minutes + 15 minutes Q&A)*
Room:	Continental
Chair:	Caitlin Mills
Talk 1:	***Your Mind on the Metaverse: What VR is Good For (And What it is Not)*** Jeremy Ballenson