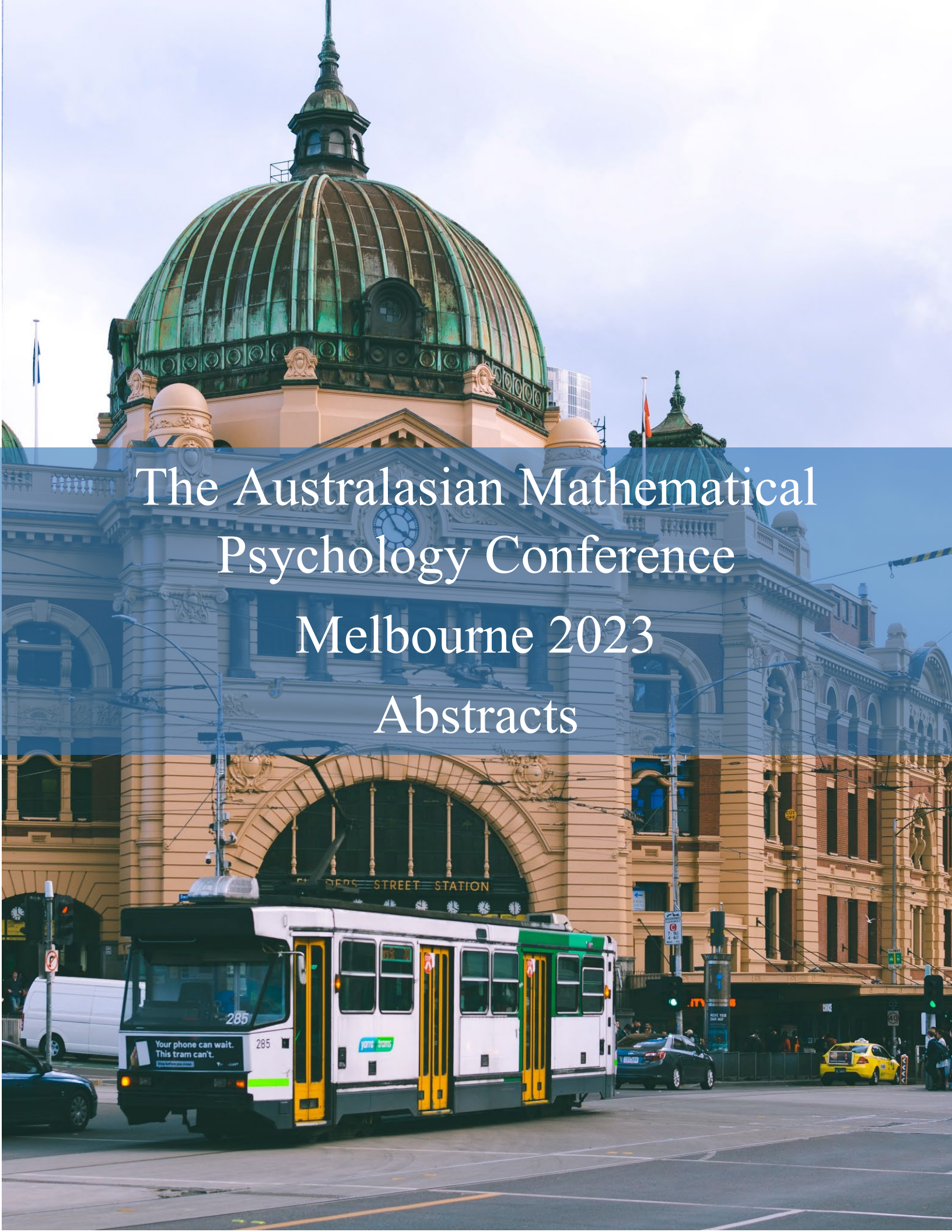


The Australasian Mathematical Psychology Conference

Melbourne 2023

Abstracts

We acknowledge the Traditional Owners of the land upon which AMPC 2023 meets, the Wurundjeri people of the Kulin nation, and pay our respects to Elders past, present, and emerging.

The University of Melbourne
February 8th - 10th 2023

Contents

Online Attendance 4

Question Time & Discord 4

Joining the server 4

Session Times 5

Wednesday 8th Feb 5

Thursday 9th Feb 6

Friday 10th Feb 7

Abstracts 8

Wednesday 8th Feb 8

Morning 1: Math & Numeracy/Visual Working Memory 8

Morning 2: Language Models 10

Afternoon 1: Language Models/Change of Mind 11

Afternoon 2: Measurement 14

Thursday 9th Feb 16

Morning Symposium 1: Human AI Teaming 16

Morning Symposium 2: Human AI Teaming 19

Afternoon 1: Memory 21

Afternoon 2: Applied Insights 23

Friday 10th Feb 26

Morning Symposium 1: Measurement/Cognitive Neuroscience 26

Morning Symposium 2: Learning 29

Events 31

Annual Football Game 31

Annual Dinner 31

Annual Business Meeting 31

ASMP Awards 31

Online Attendance

AMPC welcomes online attendees via Zoom and Discord.

Join the conference zoom using this hyperlink or by using the following room id and password:

ID: 449 067 5212
Password: 758047

Question Time & Discord

In the style of last year, to accommodate both online and in person questions, and to allow questions and conversations to occur asynchronously, AMPC has a dedicated Discord server. Discord is a free chat service that works on mobile and computer devices.

Each talk will have a dedicate room where attendees can ask questions of the speaker, discuss the presentation, meet the presenter, and share contact details.

The session chair will be able to review questions during question time. Unanswered questions can be left to the presenter to address after the talk.

You can download discord for free here: https://discord.com/download

Or you can register for discord online at https://discord.com/register and login online https://discord.com/

You can also download discord on your phone from the Android and Apple app stores.

Joining the server

After making a discord account, select the '+' to 'add a server' and select 'Join a Server'.

Copy and paste this invite link: https://discord.gg/PS6tym2jUt and click 'join server'

Once in, you will see an 'its-me' and 'initial-help' channel.

Go to **'its me'** and **type your name and affiliation** (so we don't admit spammers).

You will then be added as an 'attendee' by a conference host and have full access to the discord, including channels for each day and talk, a general discussion section, meeting rooms to talk online, and a place to leave feedback for future years.

If you can, **please join the discord server before the conference and follow the 'its me' instructions**. A conference host will admit you to the server as soon as possible, and you will be ready for the talks!

Remember, be respectful online to your fellow attendees and the presenters.

Session Times

Wednesday 8th Feb

Theme	Time	Authors	Title	Chair
Welcome	**9:00**	**AMPC 2023 Committee**	**Welcome Address**	**-**
Math & Numeracy	9:10	Gabrielle da Costa & Jian Chen	Math Anxiety Predicts Math Ability Among Adults	Paul Garrett
Math & Numeracy	9:20	Yonatan Vanunu & Roger Ratcliff	Forming Numerosity Representations: A Theory of Selective Sampling	Paul Garrett
Math & Numeracy	9:40	Roger Ratcliff & Gail McKoon	Using Diffusion Models for Symbolic Numeracy Tasks to Examine Aging Effects	Paul Garrett
VW Memory	10:00	Jason Zhou, Philip L. Smith & Adam F. Osth	The Spatiotemporal Gradient of Intrusion Errors in Continuous Outcome Source Memory	Paul Garrett
VW Memory	10:10	Paul Garrett, Ashlee King, Adam Osth, Allen Qian, & Philip Smith	Modelling retrieval in sequential visual working memory: slots, resources, or sample size?	Paul Garrett
VW Memory	10:20	Philip Smith, Saam Saber & Paul Garrett	Modelling retrieval from visual working memory using the circular diffusion model	Paul Garrett
Break	**10:40**	**-**	**Morning Tea**	**-**
Language Models	11:00	Charles Kemp, Simon Jerome Han, James Winters & Piers Kelly	Chinese characters have become more complex over time	Andrew Wang
Language Models	11:10	Simon John Dennis, Kevin Shabahang & Hyungwook Yim	A Retrieved Context Model of Sequence Generation	Andrew Wang
Language Models	11:30	Simon De Deyne & Sophie Warner	Common words, uncommon meanings: evidence for widespread gender differences in word meaning.	Andrew Wang
Break	**11:40**	**-**	**Lunch**	**-**
Language Models	1:10	Kevin Shabahang, Hyungwook Yim & Simon J Dennis	Context reinstatement as a basis for paradigmatic relations	Andrew Wang
Language Models	1:20	Andrew Wang, Simon De Deyne, Meredith McKague & Andrew Perfors	An empirical investigation of core vocabulary using word prediction in context	Andrew Wang
Change of Mind	1:30	Grant Taylor, Augustine Nguyen & Nathan Evans	Does allowing for changes of mind influence initial responses?	Brett Hayes
Change of Mind	1:40	Guangyu Zhu, Yiyun Shou & Michael Smithson	An exploratory investigation of the convergent validity and test-retest reliability of three uncertainty preference measures	Brett Hayes
Change of Mind	2:00	Brett Hayes, Saoirse Connor Desai, Joshua Pham, Jaimie Lee, Keith Ransom, & Andrew Perfors	Changing your mind about the data: Retrospective application of sampling assumptions in property inference	Brett Hayes
Change of Mind	2:10	Rebecca K West, William J Harrison , Natasha Matthews, Jason B Mattingley, & David Sewell	Common Computations Underlie Confidence Judgements in Visual and Auditory Decisions	Brett Hayes
Change of Mind	2:30	Augustine Nguyen, Guy Hawkins & Nathan Evans	Task difficulty affects response caution: Implications for interpretation (subject to change)	Brett Hayes
Break	**2:40**	**-**	**Afternoon Tea**	**-**
Measurement	3:00	Yiyun Shou, Michael Smithson & Lok Him Lee	Finite-Tailed CDF-Quantile Distributions for Modelling Subjective Certainty	Daniel Little
Measurement	3:20	Michael Smithson	The ROC Area Under the Curve (or Mean Ridit) as an Effect-Size	Daniel Little
Measurement	3:30	Daniel R. Little & Ami Eidels	Human scheduling of perceptual judgement tasks	Daniel Little
Measurement	3:50	John Dunn & Laura Anderson	The monotonic linear model:Testing for removable interactions	Daniel Little
Measurement	4:10	Chenyuan Zhang, Charles Kemp & Nir Lipovetzky	Understand and Predict Human Replan Behaviour on the Tower of London task	Daniel Little
Measurement	4:20	Jeromy Anglim, Patrick D. Dunlop, Serena Wee, Sharon Horwood, Joshua K. Wood, & Andrew Marty	Personality and Intelligence: Meta-Analysis, Item-Level Prediction, and Other-Rated Personality	Daniel Little

Thursday 9th Feb

Theme	Time	Authors	Title	Chair
Symposium 1	9:00	Tehilla Ostrovsky & Ben Newell	Natural Language Models as Tools for Validating Decision-Making Models	Garston Liang
Symposium 1	9:20	Luke Strickland, Russell Boag, Andrew Heathcote, Vanessa Bowden, Simon Farrell, Micah Wilson, Jack Hutchinson, & Shayne Loft	Understanding how humans adapt to the reliability of automated advice using cognitive models of learning and decision making	Garston Liang
Symposium 1	9:40	Russell J Boag, Luke Strickland & Shayne Loft	Automation, attention, and cognitive control in multitasking	Garston Liang
Symposium 1	10:00	Garston Liang, Guy Hawkins, Ami Eidels, Scott Brown, & Ben Newell	Is it auto or manual? The costs of seeking recommendations from an algorithm	Garston Liang
Symposium 1	10:10	Melanie McGrath, Cecile Paris, & Jessica Irons	Collaborative intelligence: the science of harnessing complementary human and machine capabilities	Garston Liang
Symposium 1	10:30	Jonathon Love, Quentin Gronau, Ami Eidels & Scott Brown	Exploring advice taking in human-AI interactions	Garston Liang
Symposium 1	10:40	Quentin F. Gronau, Jonathon Love, Ami Eidels & Scott D. Brown	A Bayesian Ideal Observer Model for Information Integration in Human-Bot Teams	Garston Liang
Break	**10:50**	**-**	**Morning Tea**	**-**
Symposium 2	11:10	Scott Brown, Quentin Gronau & Mark Steyvers	Unknown unkowns - when to step back.	Garston Liang
Symposium 2	11:20	Zheguang Zhao, Benjamin Tag, Niels van Berkel, Sunny Verma, & Benjamin Zi Hao Zhao, Shlomo Berkovsky, Dali Kaafar, Vassilis Kostakos, & Olga Ohrimenko	DDoD: Dual Denial of Decision Attacks on Human-AI Teams	Garston Liang
Symposium 2	11:30	Hector David Palada, Timothy Ballard, Son Tran, Simon Farrell, Shayne Loft, & Andrew Heathcote	Development of a computational model of adversarial decision making to support human-AI teaming	Garston Liang
Symposium 2	11:40	Murray Bennett, Laiton Hedley, Jonathon Love, Joseph Houpt, Scott Brown, & Ami Eidels*	Group performance in human-human and human-bot teams	Garston Liang
Symposium 2	11:50	Laiton Hedley, Murray Bennett, Jonathon Love, Joseph Houpt, Scott Brown, & Ami Eidels	The Relationship Between Teaming Behaviours and Joint Capacity of Hybrid Human-Machine Teams	Garston Liang
Break	**12:00**	**-**	**Lunch**	**-**
Memory	1:30	Viviana Sastre Gomez, Rebecca Defina, Paul Garrett, Jeffrey M. Zacks, & Simon Dennis	The prevalence of multitasking presents challenges for theories of event segmentation	Adam Osth
Memory	1:40	Haomin Chen & Adam Osth	Greater Target or Lure Variability? An Exploration on the Effects of Stimulus Types and Memory Paradigms.	Adam Osth
Memory	1:50	Adam F. Osth, Nicholas Bowshell, Zhenning Liu, Krista Ehinger, & Robert Nosofsky	Integrating representations of real world objects with global similarity models of recognition memory	Adam Osth
Memory	2:10	Lyulei (Raina) Zhang & Adam Osth	Modelling Orthographic Similarity Effect in Recognition Memory	Adam Osth
Memory	2:30	Sam Williams & Adam Osth	A global similarity model of false memory in the Deese-Roediger-McDermott paradigm	Adam Osth
Break	**2:40**	**-**	**Afternoon Tea**	**-**
Applied Insights	3:00	Rachel Stephens & Keith Ransom	Developing an AI-enabled Visualisation Tool to Support Reasoning from Consensus	Alex Thorpe
Applied Insights	3:10	Astrid Zeman, Tim Leers & Hans Op de Beeck	Mooney Face Image Processing in Deep Convolutional Neural Networks Compared to Humans	Alex Thorpe
Applied Insights	3:20	Gavin Cooper & Guy Hawkins	Consumer choices in the presence of multiple levels of attribute variability	Alex Thorpe
Applied Insights	3:30	Shayne Loft, Russell J. Boag, Andrew Heathcote & Hector Palada	The Application of Evidence Accumulation Modelling to Understand how Humans Adapt to Task Demands in the Modern Workplace	Alex Thorpe
Applied Insights	3:50	Zach Howard, Elizabeth Fox, Nathan Evans & Shayne Loft	Capturing the Time Dynamic Properties of Human Cognition	Alex Thorpe
Applied Insights	4:10	Alexander Thorpe, Oliver Kelly, Alex Callen, Andrea Griffin, & Scott Brown	Frog detectives: a model of citizen scientist responses to a perceptual decision-making task	Alex Thorpe
Applied Insights	4:20	Joseph Austerweil	Teaching Agents with Physical Intervention	Alex Thorpe
-	**4:40**	**-**	**Walk to Soccer Field**	**-**
-	**7:30**	**-**	**Annual Dinner**	**-**

Friday 10th Feb

Theme	Time	Authors	Title	Chair
Measurement	9:00	Micah D. Wilson, Luke Strickland & Timothy Ballard	The Journey of an Open Source Modelling Package: A Case Study of FIPS	Daniel Feuerriegel
Measurement	9:20	Niek Stevenson, Quentin F. Gronau, Reilly J. Innes, Birte U. Forstmann, & Dora Matzke	Bayesian hierarchical modelling for between-subject analysis	Daniel Feuerriegel
Measurement	9:30	Piers Howe, Andrew Perfors, Bradley Walker, Yoshihisa Kashima, & Nicolas Fay	Estimating base rate neglect and conservatism by eliciting full probability distributions	Daniel Feuerriegel
Measurement	9:50	Jess Grimmond, Guy Hawkins, Scott Brown & Reilly Innes	Modelling Self-Report Responding	Daniel Feuerriegel
Cog-Neuro	10:10	Jie Sun, Daniel Feuerriegel & Adam Osth	What causes drift rate variability in the diffusion decision model?	Daniel Feuerriegel
Cog-Neuro	10:20	Christopher Whyte, Brandon Munn, Eli Müller & James M. Shine	Apical Amplification Determines Perceptual Dominance in a Thalamocortical Model of Binocular Rivalry	Daniel Feuerriegel
Cog-Neuro	10:30	Daniel Feuerriegel, Simon Lilburn, Jane Yook, Paul Garrett, Stefan Bode, & Philip Smith	Neural correlates of decision evidence accumulation when making judgments across continuous dimensions	Daniel Feuerriegel
Break	**10:40**	-	**Morning Tea**	-
Learning	11:00	Yanjun Liu & Jennifer S. Trueblood	The impact of preference learning on context effects in multi-alternative, multi-attribute choice	Daniel Bennett
Learning	11:10	Laura Wall	Learning under changing contexts	Daniel Bennett
Learning	11:30	Steven Miletic, Niek Stevenson, Birte U Forstmann & Andrew Heathcote	Combined modelling of reinforcement learning and decision process	Daniel Bennett
Learning	11:40	Daniel Bennett & Cassandra Dods	Counterfactual learning rates in reinforcement learning are associated with affective reactivity to reward	Daniel Bennett
Learning	12:00	Manikya Alister, Keith Ransom & Andrew Perfors	Learning Pedagogical Sampling Assumptions: How do Our Assumptions About Information Providers Affect How We Learn from Examples?	Daniel Bennett
-	**12:10**	-	**Business Meeting & Awards**	-
-	**12:30**	-	**Fin**	-

Abstracts

Wednesday 8th Feb

Morning 1: Math & Numeracy/Visual Working Memory

Math Anxiety Predicts Math Ability Among Adults **9:10am**

Gabrielle da Costa & Jian Chen (Online)

Recruitment remains difficult for careers in STEM, with limited numbers of people possessing the mathematical ability required. Higher anxiety surrounding math precipitates poorer ability in math. Research reports varying relationships between math anxiety (MA) and math ability, with varying effects of gender, age, education level, and math experience. Thus, this study explored the direct predictive relationship between MA and math ability, controlling for the effects of age, gender, education level, and math experience. The results indicated that (1) higher levels of math experience predicted higher math ability scores, $B = .12$, 95% CI [.04, .19], $p < .05$, and (2) higher MA scores significantly predicted lower math ability scores, after controlling for the covariates, $B = -0.29$, 95% CI [-.43, -.15], $p < .001$. Addressing MA earlier in life may be beneficial in reducing the development of negative emotions and experiences associated with math in later life.

Forming Numerosity Representations: A Theory of Selective Sampling **9:20am**

Yonatan Vanunu & Roger Ratcliff (Online)

Two leading models of numerosity judgments suggest that numerical representations are described by a Gaussian distribution on a mental line - where the mean and SD represent the number and the variability around it. One account suggests that number and variability increase linearly, and another suggests that number increases logarithmically while variability remains systematic. Although the two accounts make similar predictions for choice with respect to Weber-Fechner's law, they can be distinguished by modeling choice and response time with the Diffusion model. However, it is unclear what drives people to use the 'linear' or the 'log' accounts, and while these models can accommodate common perceptual effects on numerosity judgments, they do not explain the cognitive process that gives rise to them. Here we present a new model that separates the two accounts according to how information is sampled selectively in response to task demands - in which perceptual properties such as screen-position (top-down) and visual-size (bottom-up) are prioritized. Moreover, by manipulating these properties in numerosity-discrimination tasks, we were able to produce performance below chance level. We also found that pupil-size modulated sampling and choice: participants with smaller pupils contracted them to form incomplete samples in high resolution - consistent with increasing variability with numerosity on a linear scale; while participants with larger pupils dilated them to perform an exhaustive sample in low resolution – consistent with systematic variability on a log scale. Thus, converging evidence from choice, response time and eye-tracking provided strong support for selective sampling as a new account that explains a battery of numerosity effects with a single cognitive mechanism.

Using Diffusion Models for Symbolic Numeracy Tasks to Examine Aging Effects **9:40am**

Roger Ratcliff & Gail McKoon

We present a model-based analysis of aging effects in three symbolic numeracy tasks. The tasks are number discrimination (is this number greater or less than 50), number memory (was this number in the list of numbers just presented), and the number-line task (point to where this number is on this number line). The first two of these tasks were fit by the standard two-choice diffusion model and the last one by the spatially continuous diffusion model (SCDM, Ratcliff, 2018). Results showed good fits of the models to accuracy and RT distributions. In the number memory task, evidence extracted from memory to drive the decision process was lower for the older adults than for young adults but for the other two (easier) tasks, there was no change in evidence with age. In all the tasks, nondecision time increased with age, but the amount of evidence needed for a decision increased in the number discrimination and number memory tasks, but not the number-line task. The number discrimination task produced conflicting accuracy and response time results as a function of age, but the model-based analyses resolved these differences. The analyses extracted differences among individuals in model components, some of which were systematic across tasks, even though the data did not allow direct comparisons.

The Spatiotemporal Gradient of Intrusion Errors in Continuous Outcome Source Memory **10:00am**

Jason Zhou, Philip L. Smith & Adam F. Osth

Previous research has characterized source retrieval as a thresholded process, which fails on a proportion of trials and leads to guessing, as opposed to a continuous process, in which response precision varies across trials but is never zero. The thresholded view of source retrieval is largely based on the observation of heavy tailed distributions of response errors, thought to reflect a large proportion of "memoryless" trials. In this study, we investigate whether these errors might instead reflect systematic intrusions from other list items which can mimic source guessing. Using the circular diffusion model of decision making, which accounts for both response errors and RTs we found that intrusions account for some, but not all, errors in a continuous-report source memory task. We found that intrusion errors were more likely to come from items studied in nearby locations and times, and were well-described by a spatiotemporal gradient model, but not from semantically or perceptually similar cues. Our findings support a thresholded view of source retrieval but suggest that previous work has overestimated the proportion of guesses which have been conflated with intrusions.

Modelling retrieval in sequential visual working memory: slots, resources, or sample size? **10:10am**

Paul Garrett, Ashlee King, Adam Osth, Allen Qian, & Philip Smith

In visual working memory, two theories have come to describe how multiple items are, or are not, stored in memory. The 'slot' model describes a limited memory model, where information beyond a limit (e.g., 4) is recalled with a precision equal to guessing. Under the 'resource' model, as the number of to-be-recalled items increases, shared memory resources deplete and precision decreases. Recent continuous report tasks and associated circular modelling have provided evidence for both the slot (Adam et al., 2017) and resource model (Oberauer, 2022), however, in our attempt to disentangle this quandary, we have come to provide a third account: a slot + sample-size model. The implications of this finding are discussed with reference to ongoing research conducted by the co-authors, regarding how set-size may play a role what were previously disparate model conclusions.

Modelling retrieval from visual working memory using the circular diffusion model **10:20am**

Philip Smith, Saam Saber & Paul Garrett

We investigated retrieval from visual working memory (VWM) in two eye-movement experiments using coloured disk stimuli that presented either a small (1, 2, 3, 4) or a large range (1, 2, 4, 6) of items across trials. We modeled retrieval from VWM using the circular diffusion model, which successfully accounted for the marginal and joint distributions of errors and response times in both experiments. Unlike previous results with perceptual tasks, the mean drift rate norm, which characterizes the strength of memory representations, was almost invariant across set sizes while the precision of the polar angle of the drift rate varied inversely with the square root of the set size. These findings are consistent with a sample-size model in which VWM representations depend on the mean of a set of noisy evidence samples that are shared among items. The shape of the drift-rate polar angle distribution was independent of its precision and invariant across set sizes and experiments, confirming and extending previous findings using two-choice decisions with grating patch stimuli. The precision of the drift-rate polar angle distribution was much lower in the large range experiment than in the small range experiment for both small and large set sizes, showing the powerful influence of contextual factors in VWM retrieval.

Morning Tea 10:40am

Morning 2: Language Models

Chinese characters have become more complex over time **11:00am**

Charles Kemp, Simon Jerome Han, James Winters & Piers Kelly

Linguistic systems are hypothesised to be shaped by pressures towards communicative efficiency that drive processes of simplification. A longstanding illustration of this idea is the claim that Chinese characters have progressively simplified over time. We tested this claim by analyzing a dataset with more than half a million images of Chinese characters spanning more than 3,000 years of recorded history. We find no consistent evidence of simplification through time, and contrary to popular belief we find that modern Chinese characters are higher in visual complexity than their earliest known counterparts. One plausible explanation for our findings is that simplicity trades off with distinctiveness, and that characters have become less simple because of pressures towards distinctiveness.

A Retrieved Context Model of Sequence Generation **11:10am**

Simon John Dennis, Kevin Shabahang & Hyungwook Yim

Transformer models of sequence generation provide a step change in the ability of connectionist models to capture a wide range of cognitive tasks. They are an existence proof that the information required to learn the control architecture and much of the representational structure of cognition is present in language input. However, as models of human cognition they have a number of undesirable properties. As a consequence of their learning rule and the reliance on hidden unit representations, they are subject to catastrophic interference not seen in human participants (McCloskey and Cohen, 1989). Furthermore, they are deep architectures that rely on the propagation of error in ways that are biologically implausible. Consequently, it is unlikely that the direct way that these architectures relate weight changes to their objective function is shared by people. To address these limitations, we present a retrieved context model (c.f. Dennis & Humphreys, 2001; Howard & Kahana, 2002) of sequence generation. Rather than have words that appear in similar contexts assume similar hidden unit patterns during learning, the model assumes that words retrieve the words from the contexts they have been seen in and that these words then retrieve paradigmatically related words at retrieval - thus eliminating slow gradient based learning and its associated problems.

Common words, uncommon meanings: evidence for widespread gender differences in word meaning. **11:30am**

Simon De Deyne & Sophie Warner

Communication relies on a shared understanding of word meaning; however, recent evidence suggests that individual variation in meaning exists even for common nouns. Consequently, understanding where and how this variation arises is integral to circumnavigating misunderstandings and facilitating more efficient communication. This study investigated the influence of gender on word meaning across two experiments. Experiment 1 used a constrained word association task where participants generated three adjectives for each of 42 words. These data were used in Experiment 2, where a separate sample judged the association strength between word pairs. Both experiments investigated the role of gender in word meaning variation and found evidence for gender-specific meaning for 11 (26%, Experiment 1) and 13 (31%, Experiment 2) out of 42 words. Experiment 2 also investigated whether conceptual diversity can be explained by gender. Using Gaussian Mixture Modelling, we found evidence for 62 clusters or concepts, with over 30% of words mapping onto multiple concepts. Evidence for gender-specific concepts was found for nearly half (46%) of the words with multiple clusters. Moreover, gender differences in meaning were not restricted to gender-stereotypical words but included neutral words as well. Altogether, the results demonstrate how speakers of the same language but with different genders have different concepts, even for common English nouns.

Lunch 11:40am

15min Fire Alarm Test between 11:30am – 12:00pm

Afternoon 1: Language Models/Change of Mind

Context reinstatement as a basis for paradigmatic relations **1:10pm**

Kevin Shabahang, Hyungwook Yim & Simon J Dennis (Online)

Words that regularly fill the same sentential slots are said to be paradigmatically related. Paradigmatic relations may be retained through a direct association at encoding or by reinstating context during retrieval. We paired proper names by embedding them into two instances of the same sentence frame, each in a separate list, yielding blocks of two study-cloze sessions. The pairing between proper names was fixed across twelve blocks. In the static condition, the same sentence frames were used across blocks, while in the dynamic condition sentence frames changed for each block. If a paradigmatic association is formed between words then interference should accrue in both conditions. If paradigmatic relations are mediated by retrieved context then changing the sentence frame should release interference. Our results are consistent with a context-reinstatement account of paradigmatic relations.

An empirical investigation of core vocabulary using word prediction in context **1:20pm**

Andrew Wang, Simon De Deyne, Meredith McKague & Andrew Perfors

Core vocabulary is a topic of huge interest in linguistics and has been studied from a wide variety of perspectives, such as language learning, dictionary studies, and cross-linguistically. In many of these conceptions, word frequency is widely considered as the conventional measure of a word's coreness; however, this approach overlooks important aspects of mental representation like centrality in an associational semantic network. In this experiment, we compare different approaches to defining core words on a task that involves predicting missing words in sentences. Results showed that core words (regardless of definition) were easier to guess than non-core words, but that frequency-defined ones did not perform as well as expected given their higher predictability and the nature of the task. Analysis on incorrect responses also showed that people preferred to guess core words, simple synonyms, and words that are taxonomically related to the target. The findings suggest that how core vocabulary is defined depends in part on the nature of the task under consideration, and that aspects of both mental representation and the linguistic environment play an important role.

Does allowing for changes of mind influence initial responses? **1:30pm**

Grant Taylor, Augustine Nguyen & Nathan Evans

Comparisons between Evidence Accumulation Models have often proved challenging due to strong mimicry in their predictions about choice response time data. One solution to reduce mimicry is constraining these models with double responses, which are a second response that is made after the initial response. However, instructing participants that they are allowed to change their mind could influence their strategy for initial responding, meaning that explicit double responding paradigms may not generalise to standard paradigms. In a recent study we have been able to provide a validation of explicit double responding paradigms by assessing whether participants' initial decisions -- as measured by diffusion model parameters -- differ based on whether or not they were instructed that they could change their response after their initial response. Across two experiments, our results consistently indicate that allowing for changes of mind does not influence initial responses, suggesting that explicit double responding paradigms should generalise to standard paradigms, and validating their use in future rapid decision-making studies.

An exploratory investigation of the convergent validity and test-retest reliability of three uncertainty preference measures **1:40pm**

Guangyu Zhu, Yiyun Shou & Michael Smithson

Studies of decision-making under uncertainty often involve measuring people's preferences for different types of uncertainty. The measurements include but are not limited to forced binary choice, certainty equivalent, and matching probability. The forced binary choice method can be applied to many scenarios that involve two choices. It is fast, but relatively lacks precision, which makes it impossible to tell who is showing a strong preference from who is showing a weak one. The certainty equivalent method, on the other hand, offers greater precision in extracting people's preferences, but its use is restricted to quantifiable outcomes. The matching probability method is a newly developed tool. It shares the advantage of the certainty equivalent method, but can also be applied to non-quantifiable outcomes. However, the convergent validity and test-retest reliability of these methods have not been well examined. The current studies investigated the convergent validity and test-retest reliability of the three methods using a gambling scenario. These studies focused on two types of uncertainty: imprecise ambiguity and conflict ambiguity. The findings revealed that preferences extracted using the three methods had weak correlations, indicating a lack of convergent validity. In addition, the test-retest reliabilities for a 9-day interval were all well below 0.7, which indicated that the preferences reflected in these methods are unstable across time. We discuss the possible reasons for the poor convergent validity and test-retest reliability as well as the implications for research in assessing uncertainty preferences.

Changing your mind about the data: Retrospective application of sampling assumptions in property inference **2:00pm**

Brett Hayes, Saoirse Connor Desai, Joshua Pham, Jaimie Lee, Keith Ransom, & Andrew Perfors

A considerable body of research suggests that, a) people adjust the inferences they draw from samples of evidence depending on their beliefs about how the samples were generated ("sampling assumptions"), and b) this process is well-captured by Bayesian models which implement different sampling assumptions by varying likelihood functions. What remains unclear, however, is whether people make such adjustments when new sampling assumptions are introduced after the sample data have been observed. Three experiments confirmed that people are capable of such retrospective adjustments. Moreover, when an initial explanation of sample generation is retracted and replaced, people adjust their inferences in line with the more recent explanation. These results suggest that people are more flexible in the way they apply their beliefs about sample generation than has been recognized previously. We discuss how such flexibility could be implemented in formal models.

Common Computations Underlie Confidence Judgements in Visual and Auditory Decisions **2:10pm**

Rebecca K West, William J Harrison , Natasha Matthews, Jason B Mattingley, & David Sewell

Humans are able to reliably evaluate their confidence in a range of different perceptual decisions. This ability is fundamental to decision-making as it allows individuals to monitor the quality of their own decisions, even in the absence of explicit feedback. Despite its importance, there is currently little consensus about the computations that humans use to evaluate their confidence and whether these computations are algorithmically related across different sensory domains. To address this question, we compared confidence judgements across two sensory modalities: vision and audition. Participants completed two versions of a categorisation task with visual or auditory stimuli. In each modality, we varied both evidence strength and sensory uncertainty. We evaluated several models from different classes which formalise the mapping of evidence strength and sensory uncertainty to confidence according to different algorithmic hypotheses. We found that a single class of models provided the best account of confidence across both the visual and auditory tasks, consistent with a common modality-independent algorithm for confidence. Our findings suggest that confidence judgements across sensory domains are best described by a single confidence algorithm whose settings are fine-tuned within each domain.

Task difficulty affects response caution: Implications for interpretation (subject to change) **2:30pm**

Augustine Nguyen, Guy Hawkins & Nathan Evans

A study by Starns & Ratcliff, 2012 has showcased the effect of task difficulty on decision threshold, where the decision threshold boundary increased with difficulty. However, as far as we are aware, no previous studies have investigated this further. The current study investigates the effect of task difficulty on subjects' decision threshold. Specifically, we programmed an experiment that cued the participants on the difficulty of the upcoming block of trials. This design allows subjects to adjust their decision threshold in accordance to the difficulty, should they choose to do so. Our experiment was able to replicate the effect of task difficulty on decision threshold. The presence of this effect has implications for the interpretation of results from between-subjects experiments (subject to change)

Afternoon Tea 2:40pm

Afternoon 2: Measurement

Finite-Tailed CDF-Quantile Distributions for Modelling Subjective Certainty **3:00pm**

Yiyun Shou, Michael Smithson & Lok Him Lee

Analyzing probability judgments that of high certainty is usually challenged by the high skewness in the distributions and values spikes at the boundaries. Most researchers either transform values at bounds into the (0, 1) interval, or apply two-stage hurdle models to separate the boundary cases from the rest. However, 0 and 1 are natural elements of probability judgments representing absolute certainty. Changing the values or assuming the process is separate from the rest of values are inappropriate. Recently, we introduced new CDF-quantile finite-tailed distributions that have defined densities at 0 and 1, and are determined by three parameters including location, dispersion and skew. In this talk, we demonstrate the application of the new distributions in modelling probability judgments involving certainty. The first example involves laypersons' interpretations of phrases that describe certainty (100%: definitely/certainly; 0%: impossible/no chance). More than half of the participants did not interpret certainty phrases as certain, and the modelling results suggested that their interpretations were significantly influenced by the decision context. The second example involves laypersons' self-reported likelihood to engage in health behaviors relating to healthcare seeking. A substantial number of participants returned likelihoods of 0% or 100%. The modelling results indicated that subjective certainty of taking health seeking behaviors was predicted by medical maximizing. We discuss the future directions of CDF-quantile finite-tailed distributions and their applications in risk communication, and research in behavioral intention, judgment and decision-making research.

The ROC Area Under the Curve (or Mean Ridit) as an Effect-Size **3:20pm**

Michael Smithson

Several authors have recommended adopting the ROC Area Under the Curve (AUC) or mean ridit as an effect-size, arguing that it measures an important and interpretable type of effect that conventional effect-size measures do not. It is base-rate insensitive, robust to outliers, and invariant under order-preserving transformations. However, applications have been limited to group comparisons, and usually just two groups, as the AUC measures the probability that a randomly chosen case from one group will score higher on the dependent variable than a randomly chosen case from the other group. I will show that the AUC can be used as an effect-size for both categorical and continuous predictors in a wide variety of general linear models (GLMs), whose dependent variables may be ordinal, interval, or ratio-level. Thus, the AUC is a very general effect-size measure and it addresses an effect not measured by conventional effect-sizes.

Human scheduling of perceptual judgement tasks **3:30pm**

Daniel R. Little & Ami Eidels

Scheduling theory concerns the development of policies determining the optimal allocation of resources to a set of tasks. Scheduling problems have been studied extensively in the context of operations research and computer science, where optimal policies have been established for many cases, but little research has examined how people perform with respect to these optimal policies despite clear applications to selective attention and behavioral allocation. We conducted several experiments in which each subtask is a random dot motion (RDM) judgment, with difficulty determined by the coherence of the motion. Participants were presented with 4 or more subtasks, which are selected one at a time and then completed. We are concerned with the order in which subtasks are selected for completion. When subtasks vary in difficulty but have the same reward value, scheduling is optimal when tasks are completed from easiest to hardest. When reward also varies, then an optimal order index can be constructed from the ratio of reward to completion time. We compared a number of manipulations varying whether tasks appeared in a consistent or varied location, the deadline for completing the four subtasks, how subtask difficulty was indicated, and the nature of the reward (e.g., game-based or points-based). We introduce novel measures of performance and a model of scheduling allowing an analysis of optimality across conditions. In general, performance is more optimal or near optimal responding when rewards are equivalent and when location is fixed and there is a deadline.

The monotonic linear model: Testing for removable interactions **3:50pm**

John Dunn & Laura Anderson

Loftus (1978, Memory & Cognition, 6, pp. 312-319) distinguished between a theoretical concept such as memory or attention, and its observed measure such as hit rate or percent correct. If the functional relationship between the concept and its measure is non-linear then only some interaction effects are interpretable. This is an example of the wider `problem of coordination' which pervades scientific measurement. Loftus drew on contributions to the representational theory of measurement (RTM) to discuss the consequences when the coordination function is assumed only to be monotonic. This led to the distinction between removable interactions that are consistent with an additive effect on the underlying theoretical concept and non-removable interactions that are not. We describe a new statistical procedure to determine if and to what extent an interaction is removable and apply it to simulated and real data sets. We focus on the possibility of recovering the form of the monotonic coordination function and the implications of this for the cognitive sciences.

Understand and Predict Human Replan Behaviour on the Tower of London task **4:10pm**

Chenyuan Zhang, Charles Kemp & Nir Lipovetzky

Understanding problem solving has been a shared challenge for both AI and cognitive science since the birth of both fields. However, the underlying mechanism of replanning behavioral receives little attention. In this work, we frame the human problem solving as Hidden Markov Model. With features extracted from AI planning approach as the input, we learn the model parameters through the human behavioral data collected in the Tower of London task and verify the effectiveness of the framework on replanning prediction in the dataset.

Personality and Intelligence: Meta-Analysis, Item-Level Prediction, and Other-Rated Personality **4:20pm**

Jeromy Anglim, Patrick D. Dunlop, Serena Wee, Sharon Horwood, Joshua K. Wood, & Andrew Marty

In this talk, I will present results from our recent meta-analysis of personality and intelligence relations (Anglim, Dunlop, Wee, Horwood, Wood, & Marty, 2022, Psych Bull) along with a large dataset Sharon Horwood and I recently collected examining the relationship between personality and intelligence using multiple other raters of personality. In particular, I will focus on the meta-analysis (N = 162,636, k = 272) of domain and facet correlations between personality and intelligence (general, fluid, and crystallized) for the major Big Five frameworks (especially NEO PI-R, Big Five Aspect Scales). I will also share results using four of our own unpublished datasets (N = 26,813) which demonstrate the ability of item-level models (i.e., 100 HEXACO personality items) to provide generalizable incremental prediction of intelligence. Finally, I will present the results of data using other-rated personality (over 1,000 focal participants who completed measures of cognitive ability each with 3 or more other ratings of personality) to assess the robustness of personality–intelligence relations. Results have implications for theorising about developmental pathways connecting intelligence and personality.

Thursday 9th Feb

Morning Symposium 1: Human AI Teaming

Natural Language Models as Tools for Validating Decision-Making Models **9:00am**

Tehilla Ostrovsky & Ben Newell

The way in which a cognitive model formalises a decision-making process may not directly mimic the underlying psychological process. Despite popular claims to the contrary (e.g., Nisbett & Wilson, 1977) participants' verbal descriptions can inform our understanding of the psychological processes underlying choice behaviour. In this talk, we will propose a novel technique that uses individuals' verbal descriptions of their problem-solving strategies to validate psychological processes assumed by decision-making models. Participants' unstructured verbal descriptions can be easily classified due to recent advances in Natural Language Processing Models (NLPM), in particular, their context-sensitivity. Across two experiments examining how people integrate social and private information when making risky decisions we compare and test NLP model outputs with individual fits of two competing cognitive models. This case study provides a preliminary, yet informative working example for future projects. We will outline the challenges in developing this approach and provide recommendations of how to overcome them. We will discuss why we think verbal descriptions provide a valuable source of data for holding decision-making models accountable to their assumptions and why we see strong potential for future applications of this method.

Understanding how humans adapt to the reliability of automated advice using cognitive models of learning and decision making **9:20am**

Luke Strickland, Russell Boag, Andrew Heathcote, Vanessa Bowden, Simon Farrell, Micah Wilson, Jack Hutchinson, & Shayne Loft

Safety-critical industries increasingly require people to incorporate automated advice in their decisions. Furthermore, automation reliability varies across technologies and contexts, and human operators adapt to different levels of automation reliability. This presentation reviews two approaches to understanding how humans work with automated advice of varying reliabilities, each with corresponding laboratory studies. The first approach applies evidence accumulation models to quantify the extent to which participants treat automation as an advisor, as compared with relying upon its advice. We found that participants primarily treated low-reliability decision aids as an advisor rather than directly accumulating evidence based on their advice. In contrast, participants more directly accumulated evidence based upon the advice of high-reliability decision aids, consistent with the role of decision aids shifting towards autonomous control of human responses. The second approach investigates variation in perceptions of the reliability of automated advice using cognitive models of learning. We found that participant reports of perceived automation reliability were adapted in a manner most consistent with a two-kernel delta rule, which specifies learning by prediction error at a rate that is sensitive to environmental volatility. However, there was substantial heterogeneity in learning processes across participants. The presentation concludes by discussing how the two approaches could be combined into an integrated model of human adaption to automation reliability.

Automation, attention, and cognitive control in multitasking **9:40am**

Russell J Boag, Luke Strickland & Shayne Loft

Many modern workplaces make use of automated decision aids to assist human operators in performing safety-critical tasks. In such settings, operator performance is driven by interdependent relationships between the operator, the automated system, and the task environment. Crucially, the ability to incorporate automated advice into decisions while performing multiple tasks depends on attention and cognitive control. Here, we propose a cognitive model that explains how individuals integrate automated advice into decision making in a cognitively demanding multitasking paradigm. The model closely fit human performance in an experiment in which we varied multitasking load (single vs. dual task) and which participants performed with and without automated assistance. Automated advice affected the decision process in two ways: by boosting the rate of processing for congruent task responses, and by inhibiting the rate of processing both for incongruent task responses and competing concurrent (non-automated) task responses. The effects of single- versus dual-task load were explained by differences in the quality and quantity of attention/processing capacity allocated across the single/dual tasks, and in differences in proactive and reactive cognitive control. These outcomes have implications for the design of decision support tools and the management of human workload.

Is it auto or manual? The costs of seeking recommendations from an algorithm **10:00am**

Garston Liang, Guy Hawkins, Ami Eidels, Scott Brown, & Ben Newell

Daily life is rife with decision tools to guide one's choices. From the mundane, such as which route to drive, to the extraordinary, such as overriding a self-driving car, algorithmic systems can provide useful recommendations to a human decision-maker. In this talk, I present dot-motion experiments that investigate 2 qualitatively distinct costs when seeking a recommendation; that of a) time when the recommendation is automatic, and b) deliberate action when the recommendation is manually solicited. To summarise our results, we find a trade-off between advice seeking and advice utilisation. When costs are low, e.g., short loading times, individuals 'seek out' the algorithm's advice more but also more readily choose against its recommendation. Larger costs, e.g., in time or deliberate effort, substantially decrease advice-seeking but result in stronger commitments to its recommendations. In this sense, increasing the costs of advice resembles commitment in the sunk-cost fallacy. We fit the Timed Racing Diffusion Model (TRDM) to the data to show that decision urgency can be a motivator to seek out a recommendation, and outline theoretical implications when algorithmic talk is cheap (ala ChatGPT) and a dominant view of advice-seeking behaviour, 'algorithm aversion'.

Collaborative intelligence: the science of harnessing complementary human and machine capabilities **10:10am**

Melanie McGrath, Cecile Paris, & Jessica Irons

Artificial intelligence (AI) is developing at a rapid pace and looks set to revolutionise many domains. While much of the focus of both research and public discussion has been on automation, it is likely that a great many activities either cannot, or should not, be fully automated. Where the task or setting is incompletely defined, or where human values and relationships are particularly important, the best results are likely to be achieved by combining human and machine intelligences. This is not simply a case of throwing together the best human and the best AI; teamwork is vital. However, there is a dearth of science to inform the design of human-AI teams. The CSIRO's organisation-wide Collaborative Intelligence (CINTEL) research program addresses this gap, bringing together researchers from psychology, machine learning/AI and applied sciences to develop the science and novel technology systems necessary to make the best use of our AI capability while empowering, rather than marginalising or substituting, human workers. Our multidisciplinary program portfolio includes foundational science such as communication and workflows for collaborative human-AI interactions, achieving shared understanding and appropriate levels of trust with AI collaborators, and the skills and mental models required for human workers to complement increasingly autonomous machines. A complementary set of projects directly apply collaborative intelligence to existing CSIRO research programs, including genome annotation, radio astronomy, cybersecurity, and marine surveillance. This presentation will introduce the CSIRO CINTEL program, share our goals, and outline our research approaches with a view to exploring the overlap with mathematical psychology.

Thursday 9th Feb – Morning 1

Exploring advice taking in human-AI interactions **10:30am**

Jonathon Love, Quentin Gronau, Ami Eidels & Scott Brown

Exploring advice taking or trust behaviour is important because of it's role in human-AI interaction the judge advisor system has been used to study advice taking behaviour under a range of conditions. The paradigm involves having participants estimate a continuous quantity, such as the price of backpacks from a list of their features, the weight of persons from their photographs, or the year of significant historic events. following their judgement, they receive an estimate from an 'advisor', and are then given the opportunity to revise their initial estimate. the extent to which participants revise their estimate toward the advisor's estimate provides a measure of 'advice taking'. We develop a variation on the task which has participants complete perceptual judgements. this allows tight control over the stimulus, and makes the task amenable to cognitive modelling. A crucial contributor to advice-taking is the perceived accuracy of the advisor. rather than providing this information "from description", our study has the participant learn the accuracy of the advisor through experience. This has many important analogues to human-AI teaming.

A Bayesian Ideal Observer Model for Information Integration in Human-Bot Teams **10:40am**

Quentin F. Gronau, Jonathon Love, Ami Eidels & Scott D. Brown

In recent years, it has become increasingly common for humans and robots to work together to achieve shared goals (i.e., human-bot teams), for instance in areas such as cybersecurity and automatic driving. Efficient collaboration requires (a) the human to incorporate information by the bot in an appropriate manner, and, ideally, (b) the bot to have a model of human behaviour so that it can optimise what information to present to the human. Here we focus as a first step on (a) and propose a model for how humans integrate bot recommendations. We present results for a simple perceptual task in which participants had to indicate the proportion of orange vs. blue squares in flashing grids in a manner similar to judge advisor systems. Our proposed model accounts for human perceptual noise, motor noise, and noise in the bot recommendation, and uses a Bayesian ideal observer approach to produce the human's final response. The model can account well for the observed data and holds the promise to not only benefit modelling of human behaviour, but to eventually also serve as a tool for optimising bot recommendations.

Morning Tea 10:50am

Morning Symposium 2: Human AI Teaming

Unknown unkowns - when to step back. **11:10am**

Scott Brown, Quentin Gronau & Mark Steyvers

Suppose one team member (human, or AI) is trying to help another, but does not know their task or objective. By observing their actions, one can infer their objective, and then try to assist. If this assistance is rebuffed, one might eventually conclude that they have no idea how to help. This conclusion is important in many settings including AI assistants, child development, statistical testing, and scientific theory development. The process by which the conclusion can be reached is not clear, both for human psychology and from an algorithmic point of view. We describe beginning steps in developing a new approach to detecting the presence of "unknown unknowns" using Bayesian inference.

DDoD: Dual Denial of Decision Attacks on Human-AI Teams **11:20am**

Zheguang Zhao, Benjamin Tag, Niels van Berkel, Sunny Verma, & Benjamin Zi Hao Zhao, Shlomo Berkovsky, Dali Kaafar, Vassilis Kostakos, & Olga Ohrimenko

Artificial Intelligence (AI) systems have been increasingly used to make decision-making processes faster, more accurate, and more efficient. However, such systems are also at constant risk of being attacked. While the majority of attacks targeting AI-based applications aim to manipulate classifiers or training data and alter the output of an AI model, recently proposed Sponge Attacks against AI models aim to impede the classifier's execution by consuming substantial resources. In this work, we propose Dual Denial of Decision (DDoD) attacks against collaborative Human-AI teams. We discuss how such attacks aim to deplete both computational and human resources, and significantly impair decision-making capabilities. We describe DDoD on human and computational resources and present potential risk scenarios in a series of exemplary domains.

Development of a computational model of adversarial decision making to support human-AI teaming **11:30am**

Hector David Palada, Timothy Ballard, Son Tran, Simon Farrell, Shayne Loft, & Andrew Heathcote

The aim of the project is to develop a computational model of adversarial decision making to support human-AI teaming. Achieving decision superiority in an adversarial context depends on the decision maker's ability to anticipate the adversary's behaviour. AI can enhance decision effectiveness by enabling humans to anticipate the adversary's likely course of action and take actions that can provide a strategic advantage. However, existing approaches to AI are typically obscured within a black box making it difficult for human to understand and trust predictions. We argue that a computational model of the adversary's behaviour is beneficial because the reasoning behind the AI can be transparent and interpretable to the human decision maker. The project is being carried out in the context of an undersea environment. Our experimental testbed includes three operators traversing an environment: A submarine operator, a high value unit (HVU) and an escort. The submarine is tasked with damaging the HVU while avoiding detection from the escort who is tasked with detecting the submarine while protecting the HVU. Our first experiment (N=46) had participants assume the role of the escort. From these observations, we developed a computational model of the escorts' search patterns. In our upcoming experiment, we will use the model to provide AI decision support to participants assuming the role of the submarine operator and examine whether decision superiority is achieved. The project provides a proof of concept of a new approach for human-AI teaming in which a model of human decision making is used as the basis for designing AI support.

Group performance in human-human and human-bot teams **11:40am**

*Murray Bennett, Laiton Hedley, Jonathon Love, Joseph Houpt, Scott Brown, & Ami Eidels**

Complex tasks may require the division of labour across multiple team members. Yet assigning multiple agents to collaborate does not guarantee efficiency. Miscommunication or limited resources may hamper the performance of the team, compared with what one might expect based on the individual performance of each operator alone. We study the performance of human-human and human-bot teams in an arcade-like computer game, Team Spirit. Two players each controlled a horizontally-moving paddle and had to prevent bouncing balls from hitting the virtual floor. Each team completed three conditions: separate, where they operated individually to maximise their own personal score while ignoring the other player; collaborative, and competitive. In another set of experiments we paired human players with a bot. Behaviour of one bot-type was driven by reinforcement learning. Another bot-type was loosely based on principles of ideal observer. Broadly, our research programme aims to scale-up cognitive modelling techniques that have been used to understand individuals' behaviour, and apply them to small groups. Here, I present measures and analyses of group performance. A follow-up talk [Hedley] will present analyses of behavioural patterns within teams, using sensitive, dynamic spatiotemporal measures of players' movement.

The Relationship Between Teaming Behaviours and Joint Capacity of Hybrid Human-Machine Teams **11:50am**

Laiton Hedley, Murray Bennett, Jonathon Love, Joseph Houpt, Scott Brown, & Ami Eidels

Artificial machine agents are becoming fully fledged autonomous team members, working alongside human co-actors to achieve outcomes neither could alone - forming the Hybrid Human-Machine (HM) team. Human-Human (HH) teams are sensitive to the social context of their environment, their behaviour will change if they are in Collaborative or Competitive contexts. As for HM teams, how their behaviour is influenced by these contexts remains unclear. Furthermore, teaming behaviours may influence the team's ability to both handle task demands and teamwork processes – what we refer to as Joint Capacity. However, global performance measures alone (such as accuracy and reaction time) cannot capture Joint Capacity or the dynamic behaviour of teams. To overcome this limitation of existing measures, we adapted the Capacity Coefficient (as a measure of Joint Capacity; Townsend & Nozawa, 1995) and state-of-the-art spatiotemporal analyses of behaviour. We compared the Joint Capacity and behavioural patterns across Team types (HH vs HM) under Collaborative and Competitive conditions. We found team behaviour predicted Joint Capacity, where less correlated behaviour was associated with better Joint Capacity, and that the behaviour of HH teams was less correlated than HM teams. Both HH and HM teams demonstrate better Joint Capacity under the Collaborative condition than the Competitive, although this benefit was much greater for HH teams than HM teams. It is not surprising that teams with two humans demonstrate superior abilities in handling both task demands and team coordination than hybrid HM teams (with only one team member capable of such abilities). Future investigations should investigate how machine agents with these abilities influence the team's behaviour and Joint Capacity.

Lunch 12:00pm

Afternoon 1: Memory

The prevalence of multitasking presents challenges for theories of event segmentation **1:30pm**

Viviana Sastre Gomez, Rebecca Defina, Paul Garrett, Jeffrey M. Zacks, & Simon Dennis

Event cognition research has typically considered events to be contiguous in time, with defined starts and ends. However, people sometimes engage in more than one event at the same time. If this happens frequently, then theories of event cognition may require modification. This research study aims to estimate how often people engage in multitasking in daily life. Ninety-seven participants were asked whether they had been multitasking at four time points during the last 24 hours. Forty-five per cent of responses reported multitasking. However, only twenty-one per cent of reports specifically listed multiple overlapping activities. Most multitasking reports showed a variety of event and duration structures. Morning, noon and night registered most event multitasking, and they were related to daily routines. The prevalence of multitasking suggests that theories of event cognition need to be expanded to accommodate non-contiguous and simultaneous events.

Greater Target or Lure Variability? An Exploration on the Effects of Stimulus Types and Memory Paradigms. **1:40pm**

Haomin Chen & Adam Osth

In recognition memory the variance of the target distribution is almost universally found to be greater than that of the lure distribution. However, these estimates commonly come from long-term memory paradigms where words are used as stimuli. Two exceptions to this rule have found evidence for greater lure variability: a short- term memory task (Yotsumoto et al., 2008) and in an eyewitness memory paradigm. In the present work, we investigated stimulus differences (faces vs. words) along with different paradigms (long-term vs. short-term) to evaluate whether either of these conditions would result in greater variability in lure items. Greater target variability was observed across stimulus types and memory paradigms. This suggests that factors other than these two might be responsible for the discrepancy in previous observations of target vs. lure variability.

Integrating representations of real world objects with global similarity models of recognition memory **1:50pm**

Adam F. Osth, Nicholas Bowshell, Zhenning Liu, Krista Ehinger, & Robert Nosofsky

The majority of models of recognition memory posit that decisions are based on the global similarity of the representation of the probe item to the representations of the stored items from the study list. If this global similarity is sufficiently high, the item is likely to be endorsed as an old item. A major limitation is that principled representations of the list items is often only possible with relatively simplistic stimuli. In this work, we generalize these models to recognition memory of real world objects by adopting representations from convolutional neural networks (CNNs) and multidimensional scaling (MDS) methods applied to similarity ratings between the objects. The results indicated that the models were very adept at capturing the hit rates, false alarm rates, and response times from two experiments manipulating category length without varying parameters between the conditions. In addition, the models were able to capture systematic variation across each category's false alarm rate.

Modelling Orthographic Similarity Effect in Recognition Memory **2:10pm**

Lyulei (Raina) Zhang & Adam Osth

Global matching models posit that recognition strength is derived from the global similarity of the probe item to each item stored in memory. However, to date many models have not adopted principled representations of word similarity. The present work is concerned with developing representations of perceptual similarity between words based on orthographic representations. Empirically, we established three key orthographic similarity effects in recognition memory, namely the replacement effect, the transposition effect, and edge effect. Subsequently, we compared a number of different orthographic representations within a global matching mode, namely slot-coding, closed-bigram, open-bigram and the Overlap model, which were compared in a hierarchical Bayesian framework. Results showed clear evidence for a replacement effect, adjacent and nonadjacent transposition effects and start letter importance in recognition memory of words. Model selection results support for the open-bigram representation.

A global similarity model of false memory in the Deese-Roediger-McDermott paradigm **2:30pm**

Sam Williams & Adam Osth

A common method for investigating false memory is the Deese-Roediger-McDermott (DRM) paradigm, in which participants are presented with lists of words related to a non-presented critical lure. Participants exhibit very high false alarm rates to the non-presented lure that sometimes even exceed the hit rate to studied items. Recently, Gatti et al (2022) demonstrated that false alarm rates in the DRM paradigm can be explained by similarity between semantic representations from the word2vec model (Mikolov et al., 2013). Specifically, they found that the similarity of the critical lure to studied words from its category (local similarity) provided a strong explanation of false recognition, which outperformed the similarity of the critical lure to all of the list items (global similarity). However, the distinction between local and global similarity was based on a linear similarity calculation that equally weights all similarity scores, which stands in contrast to established computational models which employ a non-linear transformation of similarity. In this work, we present a global similarity model using a non-linear transformation of similarity and its application to two DRM datasets. In addition, the model was applied to response times using back-end linear ballistic accumulators (LBA: Brown & Heathcote, 2008). The model was able to capture the hit rates, false alarm rates, and response time distributions across all probe types. This suggests that global similarity models remain a viable account of false recognition from the DRM paradigm.

Afternoon Tea 2:40pm

Afternoon 2: Applied Insights

Developing an AI-enabled Visualisation Tool to Support Reasoning from Consensus **3:00pm**
Rachel Stephens & Keith Ransom

As reasoners, we rarely possess the knowledge, expertise or first-hand evidence necessary to evaluate the credibility of claims critical to our daily lives. Instead, we rely on indirect evidence implicit in the information shared by others – often via the public information environment, including social media platforms. Problematically, previous research suggests that people can be misled by "false" perceptions of consensus about a given claim. Such effects may be amplified in online environments by social network structures and personalised information portals that can distort perceptions of both the level of consensus in a claim, and the quality of reasoning and evidence that support it. Our program of research aims to understand how people evaluate consensus information and incorporate cues to the independence or quality of other people's reasoning and underlying evidence. We also aim to develop and test tools such as an AI-enabled, interactive reasoning aid that can help people assess the nature and quality of consensus views gleaned via a shared information environment. This presentation will summarise the results from a series of online psychological experiments that address these aims. The experiments generally involved presenting participants with fictional social media posts about target claims, with the posts varying systematically in key factors, such as the diversity or expertise of authors and the justifications behind a reported stance. Participants were asked to rate their belief in the target claims and choose which posts they would share, either with or without the support of a visual reasoning aid. Our studies show that without support, people are often persuaded by consensus information but struggle to incorporate cues to the independence or quality of other people's recommendations and underlying evidence. However, initial findings indicate that a visual reasoning aid could successfully assist people with this struggle. This research program: 1) highlights the need to develop technologies to support people in reasoning from the information environment, and 2) has made useful headway in beginning to develop such technology.

Mooney Face Image Processing in Deep Convolutional Neural Networks Compared to Humans **3:10pm**
Astrid Zeman, Tim Leers & Hans Op de Beeck

Deep Convolutional Neural Networks (CNNs) are criticised for their reliance on local shape features and texture rather than global shape. We test whether CNNs are able to process global shape information in the absence of local shape cues and texture by testing their performance on Mooney stimuli, which are face images thresholded to binary values. More specifically, we assess whether CNNs classify these abstract stimuli as face-like, and whether they exhibit the face inversion effect (FIE), where upright stimuli are classified positively at a higher rate compared to inverted. We tested two standard networks, one (CaffeNet) trained for general object recognition and another trained specifically for facial recognition (DeepFace). We found that both networks perform perceptual completion and exhibit the FIE, which is present over all levels of specificity. By matching the false positive rate of CNNs to humans, we found that both networks performed closer to the human average (85.73% for upright, 57.25% for inverted) for both conditions (61.31% and 62.70% for upright, 48.61% and 42.26% for inverted, for CaffeNet and DeepFace respectively). Rank order correlation between CNNs and humans across individual stimuli shows a significant correlation in upright and inverted conditions, indicating a relationship in image difficulty between observers and the model. We conclude that in spite of the texture and local shape bias of CNNs, which makes their performance distinct from humans, they are still able to process object images holistically.

Consumer choices in the presence of multiple levels of attribute variability **3:20pm**

Gavin Cooper & Guy Hawkins

When assessing possible products consumers compare options on the available attribute information. This attribute information can vary in a few distinct ways. Often attribute information can be missing, such as when there are no user ratings for a product. The attribute levels can also be relatively static, with little to no variability between options. This can be seen in markets where all companies compete heavily on price. Finally, an attribute level could be highly variable. A search for food options may find 5-star restaurants displayed next to 1-star fast food outlets. To investigate the relative effects of attribute levels in these different contexts (absent, no variability and high variability) a task was designed to present participants with a sequence of choices about hypothetical restaurants. The restaurants were shown as a dollar rating (Price; 5 levels) and/or a star rating (Quality; 5 levels), and the participants were asked to respond as to whether they would consider (Accept) that restaurant or not. Participants were randomly assigned to one of 13 groups: 2 attribute absent groups (Price only and Quality only), 5 groups where the price was fixed at one of the 5 levels and quality freely varied, 5 groups where quality was fixed at one of the 5 levels and price freely varied, and the last group where both attributes varied. The groups were partitioned into two single attribute groups, the fixed price, fixed quality and freely varying groups and various statistical models were fit to each collection. The models were compared via BIC and the winning model for each partition is compared to gain insight into how the variability or absence of attribute information affects the processing of the other attributes. Vastly different response patterns were evident in the single attribute cases, and more subtle differences will be shown between the double attribute groups.

The Application of Evidence Accumulation Modelling to Understand how Humans Adapt to Task Demands in the Modern Workplace **3:30pm**

Shayne Loft, Russell J. Boag, Andrew Heathcote & Hector Palada

In this presentation I will argue for wider adoption of cognitive modelling in Human Factors research. Evidence accumulation models are a class of computational cognitive model used to understand the latent cognitive processes that underlie human decisions and response times. They have seen widespread application in cognitive psychology and neuroscience. However, historically, their application was limited to simple decision tasks. I will overview research applying evidence accumulation models to gain insight into the cognitive processes that underlie human behaviour in simulations of applied domains, such as air-traffic control (ATC), driving, forensic and medical image discrimination, and maritime surveillance. The outcomes have provided unique insights into how the human cognitive system adapts to task demands and interventions, such as task automation, and can be used by practitioners to proactively identify red zones of workload (overload), understand the nature of expertise, improve training and work design and to inform the design of automated decision-support tools. Basic researchers also benefit from understanding how cognitive theories generalise to representative work tasks. Future research directions and challenges, such as integrating evidence accumulation models with learning models and broader cognitive architectures, and associated opportunities for research collaboration, will be outlined.

Capturing the Time Dynamic Properties of Human Cognition **3:50pm**

Zach Howard, Elizabeth Fox, Nathan Evans & Shayne Loft

Despite the ubiquitous nature of Evidence Accumulation Models in cognitive and experimental psychology, there has been a comparatively limited uptake of such techniques in the applied literature. While quantifying latent cognitive processing properties has significant potential for applied domains such as adaptive work systems, accumulator models often fall short in practical applications. Two primary reasons for these shortcomings are the complexities and time needed for the application of cognitive models, and the failure of current models to capture systematic trial-to-trial variability in parameters. We have developed a novel, trial-varying extension of the Shifted Wald model to address these concerns. By leveraging conjugate properties of the Wald distribution we derive analytic solutions for threshold and drift parameters which can be updated instantaneously with new data at a single-trial time scale. The resulting model allows the quantification of systematic variation in latent cognitive parameters across trials and we demonstrate the utility of such analyses through simulations and an exemplar application to an existing data set. The analytic nature of our solutions opens the door for real-world application, significantly extending the reach of computational models of behavioral responses.

Frog detectives: a model of citizen scientist responses to a perceptual decision-making task **4:10pm**

Alexander Thorpe, Oliver Kelly, Alex Callen, Andrea Griffin, & Scott Brown

Crowdsourced data analysis represents an attractive solution for researchers working with large datasets that must be analysed by a human, such as classifying noisy audio recordings. It also allows members of the public without formal training to be involved in the scientific process, by volunteering as citizen scientists. However, there is the potential for inexpert data analysis to be incomplete and inaccurate, and for inaccuracies in large datasets to go uncorrected, which would undermine the utility of crowdsourced analysis. To address these issues, we applied a model based on cultural consensus theory to a large dataset of audio field recordings taken from a frog breeding pool and analysed by citizen scientists. Volunteer participants were presented with noisy brief recordings and asked to identify which species of frog, if any, were calling on the recordings. 358 participants were asked to identify target species of ground-dwelling frogs, while 309 participants were asked to identify target species of tree frogs. 23 participants completed both tasks. An expert researcher also evaluated each of the 1260 recordings, to establish a "ground truth" for each recording. Through modelling, the probability of each clip containing each species of frog was estimated, along with the competence of bias towards a "yes" response of each participant. Group consensus of which recordings were likely to contain a target frog agreed with the expert analysis, with some evidence that the model-based consensus performed well even when estimated from a sparse dataset, where most recordings had not been analysed.

Teaching Agents with Physical Intervention **4:20pm**

Joseph Austerweil

Over the last decade, intelligent machines have become an indispensable part of daily life. However, this is typically only successful when humans learn to interact with the machines, rather than machines learning to interact with humans. Recent work has found that machines can be more natural to interact with when machine agents are endowed with appropriate psychological biases and knowledge. In this talk, I will present a task domain that is relatively unexamined: teaching agents by physical intervention. I will discuss how different ways of interpreting human physical state changes (i.e., moving the agent) affect how successfully a machine agent can learn from the intervention.

Walk to Football Field 4:20pm

Annual Dinner 7:30pm

Friday 10th Feb

Morning Symposium 1: Measurement/Cognitive Neuroscience

The Journey of an Open Source Modelling Package: A Case Study of FIPS **9:00am**

Micah D. Wilson, Luke Strickland & Timothy Ballard

Most mathematical models in the human sciences are implemented programmatically in code, but there exists great variability with respect to the reproducibility and accessibility of modelling software. There exist strong arguments that going beyond a minimum state of reproducibility (e.g., providing only formulae), towards a gold standard where flexible model implementations are provided with journal articles, can help foster robust and cumulative science and mitigate research fragmentation. In this presentation, we focus on this problem as it applies in the domain of biomathematical models (BMMs) — a class of phenomenological parametric models which are used to predict fatigue-related risk across many industries. We detail the motivation for and developmental history of FIPS, an open-source R package that implements several BMMs. Using FIPS as an illustrative case-study, we will argue that there are major benefits to investing the resources required to transform ad-hoc modelling code, to well documented, accessible software. The talk will also include some discussion of the challenges specific to biomathematical modelling, and outline the future directions for the FIPS package, and how other researchers can leverage this tool.

Bayesian hierarchical modelling for between-subject analysis **9:20am**

Niek Stevenson, Quentin F. Gronau, Reilly J. Innes, Birte U. Forstmann, & Dora Matzke

Cognitive models are more and more frequently applied to test both within- and between-subject hypotheses, however, the latter has generally suffered from the lack of statistical methods to answer such questions. A common approach is to perform a second step of analysis on the estimated parameters of the model to answer whether, for example, drift rate differs with age, or between people with schizophrenia and controls. However, a lot of statistical power is lost in such two-step analyses. Recently, Boehm et al. (2021) proposed to include linear models such as ANOVA, regression, and by extension mixed effect models, in the hierarchical framework in which cognitive models are usually estimated. With such a hierarchical linear model we omit the two-step analysis. Furthermore, we can calculate Bayes factors between the null and the proposed model. In the current work, we generalize these results to a user-friendly package with multiple types of cognitive models. Lastly, we explore additional levels of hierarchy that govern different groups. Our work provides researchers with the option to formalize different types of hypotheses for between-subject research, whilst also maintaining a more parsimonious parameter space. The method is built on the robust and fast particle metropolis within Gibbs sampling algorithm (Gunawan, Hawkins, Tran, Kohn, & Brown, 2020). Furthermore, for all methods we also developed unbiased marginal likelihood estimation based on importance sampling squared (Tran, Scharth, Gunawan, Kohn, Brown, & Hawkins, 2021). We provide open source code and accompanying tutorial documentation to make the method accessible to researchers.

Estimating base rate neglect and conservatism by eliciting full probability distributions **9:30am**

Piers Howe, Andrew Perfors, Bradley Walker, Yoshihisa Kashima, & Nicolas Fay

Bayesian statistics provides a theoretical framework for understanding how people update their beliefs (i.e., their priors) in light of new evidence (i.e., the likelihood). Previous research has indicated that people tend to underweight the prior information (known as base rate neglect) and the likelihood (referred to as conservatism) when making probability estimates. However, this work has relied on single-point estimates of probability (e.g., “What is the probability that A occurs?”). In the present study, we elicited and fit full probability distributions for each individual. This allowed us to measure, for the first time, the degree of base-rate neglect and conservatism simultaneously at the level of the individual. We found that while most participants disregarded the base rate, they did so less when the prior was made explicit. Many participants were also conservative, but we observed no systematic relationship between base rate neglect and conservatism within individuals. Overall, our method shows promise in that it allows us to address questions that cannot be addressed by eliciting single-point probability estimates.

Modelling Self-Report Responding **9:50am**

Jess Grimmond, Guy Hawkins, Scott Brown & Reilly Innes

Self-report measures are widely used in the psychological sciences as they can seemingly assess latent aspects of people's personality, cognition and behaviours which are otherwise difficult to capture. However, self-report tasks are often measured on scales which require subjective interpretation and are open to biased and deceptive responding. Hence, it is unclear whether responses to these tasks accurately capture the information they aim to. Furthermore, scores from self-report tasks rarely correlate to performance on behavioural tasks which supposedly measure the same constructs (Dang, King and Inzlict, 2021), which has led some to question their continued and wide-spread use. Our research aims to develop a model of survey responding which reduces the noise from individual differences in response style and aims to improve the precision of information gain. Modelling self-report responding in this way may provide a deeper understanding of self-report behaviour and facilitate more accurate analyses of both survey and behavioural tasks when modelled jointly. This work is in the early stages of conception and this presentation will provide an overview of the issue and our suggested approach to modelling self-report data with the aim of generating discussion and feedback.

What causes drift rate variability in the diffusion decision model? **10:10am**

Jie Sun, Daniel Feuerriegel & Adam Osth

The diffusion decision model (DDM) has been successfully accounting for reaction times distribution and accuracy for simple choice tasks. However, the model has been criticized for its ad hoc distributional assumptions of cross-trial variability in its parameters, such as the drift rate, which is responsible for the prediction of slow errors. It's interesting to see how the drift rate variability estimates would change if the trial-by-trial variability is accounted for. Using a recognition memory task with electroencephalography recording, the current study aims to include systematic variances in drift rate, from both exogenous factors (e.g., word frequency and study-test lag) and EEG signals corresponding to retrieval success as endogenous factors, to account for random trial-to-trial variability in drift rate. Our results have shown when assuming individual trial drift rate as a linear combination of these factors, better model fits were observed with more factors are included, indicated by superior model selection scores. We have also observed greater systematic variabilities with more factors included in the model. However, we did not observe large corresponding decreases in the random drift rate variability estimates compared to a standard model without any factors included. This suggests that the random drift rate distribution may be impossible to be fully accounted, and that other mechanisms might be involved in producing slow errors which are not implemented in DDM.

Apical Amplification Determines Perceptual Dominance in a Thalamocortical Model of Binocular Rivalry **10:20am**

Christopher Whyte, Brandon Munn, Eli Müller & James M. Shine

Uncovering the neural mechanisms of consciousness remains one of the great challenges in modern neuroscience. Most of our current understanding is derived from human cognitive neuroscience where high resolution recordings and precise causal manipulations are exceedingly difficult, if not impossible, limiting the neurobiological detail of both theory and modelling. Recently, however, great progress has been made in a mouse model of conscious perception where a recurrent circuit connecting layer 5b pyramidal neurons to non-specific matrix thalamus has been identified as causally relevant to conscious perception in a simple threshold detection task. Here we use computational modelling to bridge the gap between the cellular mechanisms of consciousness identified in the mouse model to the more sophisticated tasks used to study conscious perception in humans. Specifically, we show that the same mechanism of apical amplification that underlies threshold detection in rodents determines perceptual dominance in a thalamocortical spiking neural network model of binocular rivalry – a staple task in the cognitive neuroscience of consciousness. The model explains, in cellular terms, the lawlike relation between stimulus properties and dominance durations (Levelt's propositions), and provides a set of novel predictions about the cellular basis of binocular rivalry and the mesoscale neural correlates or rivalry.

Neural correlates of decision evidence accumulation when making judgments across continuous dimensions **10:30am**

Daniel Feuerriegel, Simon Lilburn, Jane Yook, Paul Garrett, Stefan Bode, & Philip Smith

The process by which we transform sensation into action is known as perceptual decision-making. This capacity has been predominantly investigated using discrete choice tasks, in which observers choose between multiple, distinct options. We can also make judgments along continuous dimensions, such as determining the exact colour of a paint swatch or the movement direction of a ball. We recently developed a Circular Diffusion model that accounts for joint distributions of accuracy and response times in continuous report tasks. In the current study we mapped the event-related potential (ERP) correlates of decision evidence accumulation as specified in this model. Participants viewed briefly-presented colour patches and reported the colour of that patch on a colour wheel using ballistic mouse movements. We identified an ERP component over parietal electrodes that closely resembled evidence accumulation trajectories in the Circular Diffusion model. Our findings establish clear neural correlates of evidence accumulation in the Circular Diffusion model, and pave the way for joint modelling of behavioural and neural data in continuous report tasks.

Morning Tea 10:40

Morning Symposium 2: Learning

The impact of preference learning on context effects in multi-alternative, multi-attribute choice **11:00am**

Yanjun Liu & Jennifer S. Trueblood

Prior experience with options can impact our preferences and future choices. When strong preferences exist, context effects are hypothesized to diminish (Huber et al., 2014). In this study, we probed the effect of prior experience with options on the strength of attraction and compromise effects. Participants had the opportunity to choose from simple cognitive tasks (i.e., counting jobs) and complete the selected task. Using an ecologically valid and incentivized task, we found evidence for the formation of strong preferences with experience, yet reversals of context effects did not attenuate. The results were replicated in a pre-registered study and showed that our findings are robust to payment schemes and display format. These findings suggest that relative evaluation still plays a role in human decision-making, even when inherent preferences are accessible. We suspect what was learned from experience in our tasks is the weights for various attributes. As predicted by many models of multi-alternative, multi-attribute choice, context effects can emerge with unequal attribute weights formed through, for instance, prior experience with options.

Learning under changing contexts **11:10am**

Laura Wall

Predictive inference is an important cognitive function and there are many tasks which measure it, and the error driven learning that underpins it. Context is a key contribution to this learning, with different contexts requiring different learning strategies. A factor not often considered however, is the conditions and time-frame over which a model of that context is developed. This study required participants to learn under two changing, unsignalled contexts with opposing optimal responses to large errors - change-points and oddballs. The changes in context occurred under two task structures: 1) a fixed task structure, with consecutive, short blocks of each context, and 2) a random task structure, with the context randomly selected for each new block. Through this design we examined the conditions under which learning contexts can be differentiated from each other, and the time-frame over which that learning occurs. We found that participants responded in accordance with the optimal strategy for each contexts, and did so within a short period of time, over very few meaningful errors. We further found that the responses became more optimal throughout the experiment, but only for periods of context consistency (the fixed task structure), and if the first experienced context involved meaningful errors. These results show that people will continue to refine their model of the environment across multiple trials and blocks, leading to more context-appropriate responding - but only in certain conditions. This highlights the importance of considering the task structure, and the time-frames of model development those patterns may encourage. This has implications for interpreting differences in learning across different contexts.

Combined modelling of reinforcement learning and decision process **11:30am**

Steven Miletic, Niek Stevenson, Birte U Forstmann & Andrew Heathcote

Learning and decision-making are interactive processes, yet cognitive modeling of error-driven learning and decision-making have largely evolved separately. Recently, evidence accumulation models (EAMs) of decision-making and reinforcement learning (RL) models of error-driven learning have been combined into joint RL-EAMs that can in principle address these interactions. We compare the most commonly used combination, based on the diffusion decision model (DDM) for binary choice, and a new RL-EAM, based on an advantage racing diffusion (ARD) framework, to test how well they can explain response time distributions, choice, and learning-related changes in response times and choices. We show that the RL-EAM framework can capture these effects as well as stimulus difficulty, speed-accuracy trade-off, stimulus-response-mapping reversal effects, and extends beyond binary choice.

Counterfactual learning rates in reinforcement learning are associated with affective reactivity to reward **11:40am**

Daniel Bennett & Cassandra Dods

Reinforcement learning is a paradigm concerned with understanding how agents learn and adjust their behaviour in response to reward and punishment. However, rewarding and punishing stimuli do not only affect learning and behaviour: they also reliably produce emotional reactions. An important open question is what role these emotional reactions play in the learning process itself. In the present study, we approached this question from an individual-differences perspective, asking whether individuals who show stronger emotional responses to rewarding and non-rewarding feedback also learn more from that feedback. We recruited 121 participants via Prolific to complete a simple two-armed bandit task in which we embedded high-resolution sampling of participants' subjective emotional valence. Within the task, we also manipulated on a trial-by-trial basis whether participants were provided with factual feedback (outcome of chosen option), counterfactual feedback (outcome of unchosen option), or both. We used two separate hierarchical Bayesian computational models to investigate participants' reinforcement learning, on the one hand, and their emotional reactivity, on the other. When we investigated correlations between participant-level parameters from the two models, we found a significant association between the rate at which participants learned from counterfactual feedback and their emotional reactivity to counterfactual feedback ($r = .3$, $p < .01$). More unexpectedly, we also found that emotional reactivity to counterfactual feedback was associated with the rate at which participants learned from factual feedback ($r = .27$, $p < .01$). Our results suggest that higher emotional reactivity might be a trait-level phenotype that is associated with faster learning from counterfactual information.

Learning Pedagogical Sampling Assumptions: How do Our Assumptions About Information Providers Affect How We Learn from Examples? **12:00pm**

Manikya Alister, Keith Ransom & Andrew Perfors

Sampling assumptions --- the assumptions people make about how an example of a category or concept has been chosen --- help us learn from examples efficiently. One context where sampling assumptions are particularly important is in social contexts, where a learner needs to infer the knowledge and intentions of the information provider and vice-versa. The pedagogical sampling assumptions model describes a Bayesian account of how learners and providers should behave given different assumptions they have about the other (e.g., is the provider trying to deceive or help me? Does the learner trust me?). In this study we tested how well this model could describe learning behaviour in the rectangle game, where a fictional information provider revealed clues about the structure of a rectangle that the learner (a participant) needed to guess. Participants received clues from either a helpful information provider, a random provider, or one of two kinds of a deceptive providers. We found that people efficiently and in line with model predictions learned the strategy of the helpful information provider, but struggled to learn the strategies of the deceptive providers.

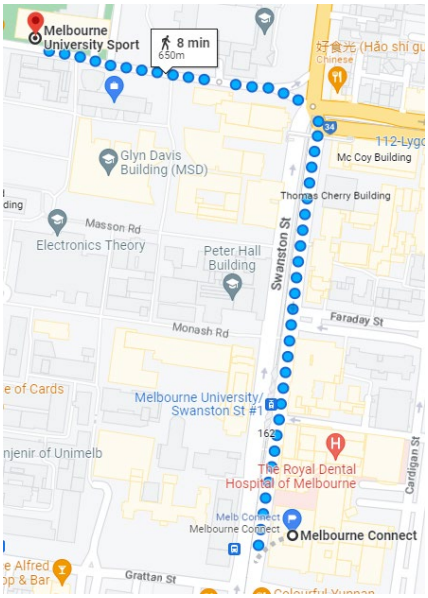
Business Meeting & ASMP Awards 12:10pm

Conference Ends 12:30

Events

Annual Football Game

The annual football game will be held on Thursday Feb 9th 2023 from 5pm – 6:30pm at the Melbourne University Sports field (see map for directions from Melbourne Connect).



Annual Dinner

This year the annual conference dinner will be held after the football match, Feb 9th 2023 from 7:30pm at the Italian restaurant, Kaprica (50 Gratton St, Carlton).

Tickets are $90 and can be purchases as part of conference registration via Eventbrite.

We encourage all attendees to join this event.

Annual Business Meeting

The annual business meeting will occur at 12:10pm on Friday 10th February 2023. All attendees are welcome.

ASMP Awards

The Australasian Society for Mathematical Psychology awards will be awarded as part of the business meeting.

If you experience any issues during the conference, please see the conference organizers:

Paul Garret, Andrew Perfors, Adam Osth, Daniel Bennett,

Daniel Little, Sasha Hart, and Viviana Sastre

or email ampccommittee@gmail.com