2024 EMPG Meeting

European
Mathematical Psychology
Group

EMPG 2024
Tübingen

Program and Abstracts

University of Tübingen, September 04 – 06, 2024

Version: 04 September 2024

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Welcome

Dear Colleagues,

It is our great pleasure to welcome you to the 2024 Meeting of the European Mathematical Psychology Group held at the University of Tübingen, Germany. We hope that this 53rd EMPG Meeting will be stimulating and productive, and that you will enjoy the ambience of the city of Tübingen and the conference venue "Alte Frauenklinik", a historical building from 1890 (formerly the University Women's Hospital) that now hosts the Department of Psychology.

This year's conference features keynote addresses from Francis Tuerlincks, Clintin P. Davis-Stober, and Julia Haaf as well as an invited symposium in memoriam of Allen Parducci.

We would like to acknowledge the generous financial support of

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

(grant HE 1771/9-1). Their support made an important contribution to many aspects of the conference, including the keynote talks and the organization of the invited symposium.

The organizing committee

Jürgen Heller
Florian Wickelmaier
Katharina Naumann
Julian Mollenhauer
Dorina Kohler
Alice Maurer

General Information

Conference Organization

Scientific Committee

Rocío Alcalá-Quintana (Complutense University of Madrid, Spain)
Hans Colonius (University of Oldenburg, Germany)
Adele Diederich (University of Oldenburg, Germany)
Jean-Paul Doignon (Université Libre de Bruxelles, Belgium)
Luca Stefanutti (University of Padova, Italy)

Organizing Committee

Jürgen Heller
Florian Wickelmaier
Katharina Naumann
Julian Mollenhauer
Dorina Kohler
Alice Maurer

Venue

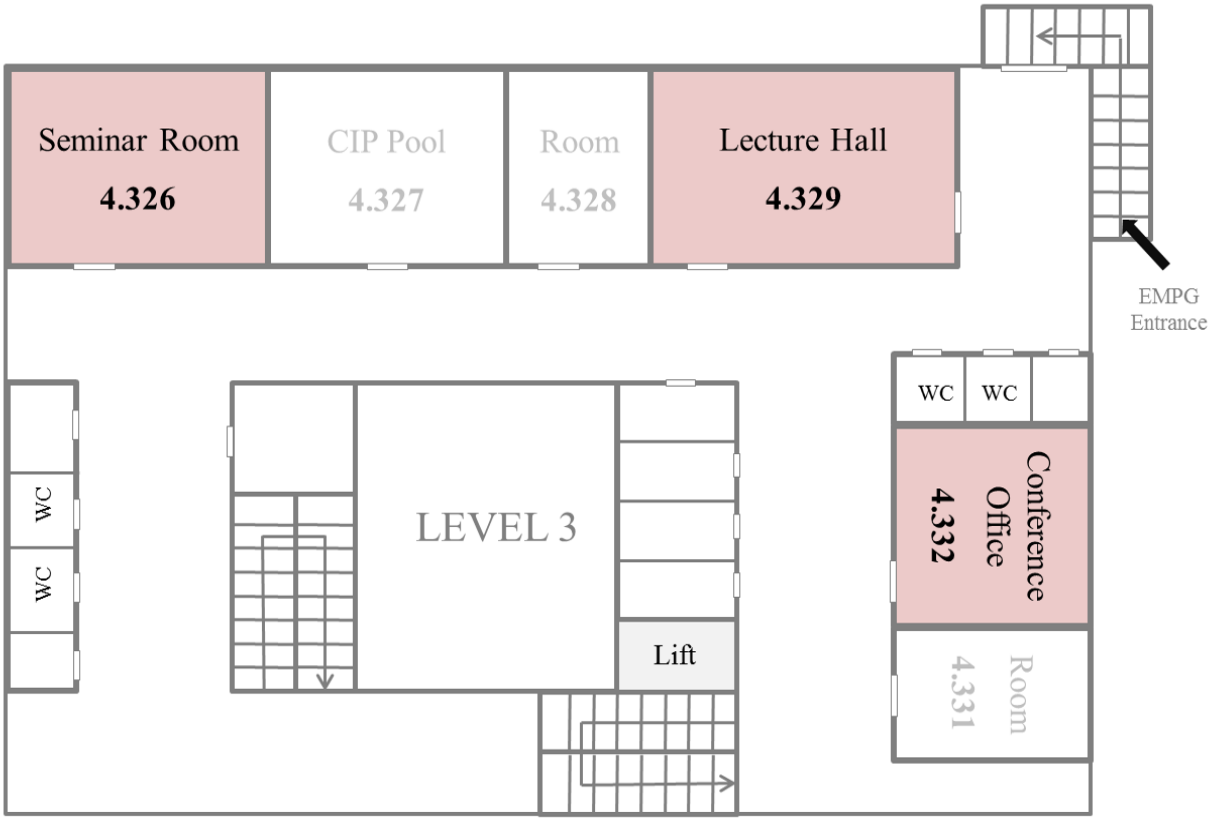
All sessions will take place in the building "Alte Frauenklinik", where the Department of Psychology is located.

Address:
Department of Psychology
University of Tübingen
Schleichstr. 4
72076 Tübingen (Germany)

All events take place on Level 3. Handicapped persons please use the Level 2 entrance and the elevator to Level 3 to access the conference rooms:

- Conference Office (4.332): Registration, Information Desk, Coffee Breaks
- Lecture Hall (4.329): All Talks

- Seminar Room (4.326): Poster Session



Conference Office

During the conference, the registration and information desk can be found in the conference office in room 4.332. Opening hours: 08:15 to 18:00. On Tuesday, September 03, the conference office will be open from 18:00 to 20:00 during the Welcome Reception.

Welcome Reception

The Welcome Reception will be on Tuesday, September 03, in and around the conference office from 18:00 to 20:00. There will be some refreshments and the opportunity to meet other conference participants.

Internet Access

You can access the internet access via the SSIDs `eduroam` and `UTguest` (no authentification).

Lunch and Coffee Breaks

Refreshments and snacks will be provided during coffee breaks on Wednesday, Thursday, and Friday in room 4.332 (conference office).

Tübingen offers many lunch options for trying local Swabian dishes or enjoying international cuisine. Most of the places are located in the old town and within walking distance from the conference venue. Important locations including hotels and a selection of restaurants can be found on the following Google Map.

Conference Dinner

The conference dinner will take place on Thursday, September 05, from 19:00 to 22:00 at the restaurant "Neckawa" (Wöhrdstr. 25, 72072 Tübingen), situated directly at the river Neckar (Google Map).

Travel

IMPORTANT: During the EMPG 2024 there is no train connection between Stuttgart main station and Tübingen. There is a bus service instead. Buses are leaving in front of Stuttgart main station (stop "Ersatzhaltestelle Kriegsbergstraße"). Follow the signs "Bus IRE 6".

Tübingen can also be reached by plane and bus. The fastest connection between Stuttgart Airport and Tübingen train station (Hauptbahnhof) is the so-called Airport Sprinter (Bus Line No. 828) or Bus Line No. X82.

The venue can be reached via a 20 minutes walk from the main train station, or by taking one of the busses at platform B (Lines No. 13, 18 or 19, exit at stop "Hölderlinstraße", Line No. 5 exit at stop "Uni Kliniken Tal"). For more information see here (in German).

There are various parking garages close to the venue (for a map see here).

Accommodation

The following hotels offered special rates beforehand:

Hotel Meteora
Weizsäcker Strasse 1
72074 Tübingen
Phone: +49 7071 970 90 20
E-mail: info@hotel-meteora.de

Hotel Domizil
Wöhrdstrasse 5-9
72072 Tübingen
Phone: +49 7071 139 0
E-mail: info@hotel-domizil.de

ibis Styles Tübingen
Friedrichstrasse 20
72072 Tübingen
Phone: +49 7071 75880
E-mail: smile@ibisstyles-tuebingen.com

Presentation Guidelines

Talks

Talks are scheduled for 20 min, including 5 min for discussion. We kindly ask speakers and chairs to adhere to this time schedule.

The lecture hall is equipped with a Windows computer for both MS PowerPoint and PDF presentations (while the latter is preferred). In case of a PowerPoint presentation we recommend to use standard fonts to minimize the risk of layout distortions.
Please notice that it is not possible to use your own notebook to avoid compatibility issues and to save time. Please make sure to transfer your presentation files to the computer in the lecture hall well before the session starts using a USB memory stick (preferably already in the morning or during the break preceding the session). Ask the technical assistants in the lecture hall for help.

Posters

The poster session is on Wednesday, September 04, from 15:40 to 17:00 in seminar room 4.326. Please attach your poster to the board (indicated by your name) until the end of the lunch break at 14:00. The poster boards allow for paper size A0 in portrait format.

Program & Abstracts

Wednesday, 04 September

08:30 - 08:50		Opening	
08:50 - 09:10	Measurement	A Theory of Psychological Variables and their Measurement Based on the Selection of Relational Responses	Matthias Borgstede, Frank Eggert
09:10 - 09:30		The Order Multiverse of Sum Scores	Luca Stefanutti, Marina Ottavia Epifania, Debora de Chiusole, Pasquale Anselmi
09:30 - 09:50		Some Results on Random Utility Theory	Hans Colonius
09:50 - 10:10		Coffee Break	
10:10 - 10:30	Response Times	Drift Diffusion Model explains How Affective Task Content Modulates Switch Costs regarding Cue-to-Stimulus Interval in Task Switching	Maryam Sadeghi Talarposhti, Leif E. Langsdorf, Torsten Schubert
10:30 - 10:50		How to (Really) Measure the Speed of Top-down Stopping Signals	Bompas, Aline
10:50 - 11:10		Interaction Contrasts for Choice Responses	Matthias Gondan
11:10 - 11:30		Variable, Sometimes Absent, but Never Negative: Applying Multilevel Models of Variability to the Backward Crosstalk Effect to Find Theoretical Constraints	Christoph Naefgen, Robert Gaschler
11:30 - 12:30	Keynote	Geometry of Statistical Models	Francis Tuerlinckx
12:30 - 14:00		Lunch Break	
14:00 - 14:20	Attention, Perception & Psychophysics	Can the Queueing Model of Visual Search Account for Feature Search?	Yiqi Li
14:20 - 14:40		On the Nature of Cross-Modal Correspondences	Jürgen Heller
14:40 - 15:00		Linear versus Logarithmic Representations of Time in Duration Discrimination	Miguel A. García-Pérez, Rocío Alcalá-Quintana
15:00 - 15:20		A Unified Approach to Understanding Order Effects in Sensory Discrimination	Rocío Alcalá-Quintana, Miguel A. García-Pérez
15:20 - 15:40		Coffee Break	
15:40 - 17:00		Poster Session	
17:00 - 18:00		Young Scientists Networking Event	

Session: Measurement

Chair: H. Colonius

A Theory of Psychological Variables and their Measurement Based on the Selection of Relational Responses

Matthias Borgstede, Frank Eggert

Otto-Friedrich Universität Bamberg

We present a theory-based approach to psychological measurement based on the idea that psychological variables are emergent properties of relational responding. Empirical findings suggest that relational responding is operant behavior and, as such, selected by its environmental consequences. Following the formal requirements of representational measurement, we propose that a measurable psychological variable emerges, if and only if relational responses are selected such that the resulting structure is isomorphic to a subset of the real numbers. Given such an isomorphism, psychological variables can be measured either directly by observation of the relational response structure, or indirectly by means of a consistent set-mapping between the objects of the relational response structure and a corresponding numerical structure (e.g., a rating scale). We further embed the concept of emergent psychological variables in the theory of operant selection, where selection is understood as a multilevel process that shapes behavior with regard to its expected fitness consequences. Since fitness itself has a quantitative structure, quantitative psychological variables should emerge if and only if the environmental contingencies allow for a reliable prediction of expected amount of fitness gain. Consequently, measurable psychological variables are context-dependent fitness predictors that may function as discriminative stimuli for expected fitness gains given a certain action is taken in that context. The theory clarifies what psychological variables are, why they are sometimes quantitative and sometimes not, how they can be measured and why we should expect quantity in the realm of psychology in the first place. Moreover, the theory is empirically rich, yielding several unique hypotheses that can be tested experimentally.

The Order Multiverse of Sum Scores

Luca Stefanutti, Marina Ottavia Epifania, Debora de Chiusole, Pasquale Anselmi

Università di Padova

The sum score (SS) on a psychological test is central in many psychometric approaches, even with ordinal data. However, comparisons involving sums of ordinal data are not meaningful, in general. Let V be a totally ordered set of $m > 1$ response categories, and $\mathcal{F}$ be the family of all the monotone strictly increasing mappings from V to the reals. With n items, V^n is the collection of all possible response patterns, and each mapping $\phi \in \mathcal{F}$ is associated with a weak order $\precsim_\phi$ on the set V^n such that, for all pairs of response patterns $x, y \in V^n$, $x \precsim_\phi y$ iff $\sum_i \phi(x_i) \leq \sum_i \phi(y_i)$. Then, the order multiverse of sum scores for the pair (m, n) of positive integers is the collection $\mathcal{M}(m, n)$ of all possible weak orders $\precsim_\phi$ on V^n for any $\phi \in \mathcal{F}$. With $n > 1$ polytomous ($m > 2$) items, the order multiverse is not singleton, and its intersection $R = \bigcap \mathcal{M}(m, n)$ is a quasi-order. Two response

patterns x and y are comparable with respect to R if and only if the statement "SS of x less or equal to SS of y" is (quantitatively) meaningful. We characterize the quasi-order R and establish the necessary and sufficient conditions for two response patterns being comparable. A rather counter-intuitive result shows that the proportion of meaningful comparisons decreases as the number of items and the number of response categories increase. It follows that alternative plausible choices for assigning reals to the response categories in V lead to a multiverse of results. Some examples with empirical data of partial credit items from the tower of London test show the paradoxical results that one obtains in attempting to order individuals by using SS produced by alternative scoring methods.

Some results on random utility theory

Hans Colonius

Carl von Ossietzky Universität Oldenburg

I present some new (or forgotten) results on certain classes of random utility models for best-choices, in particular models with stochastic dependency. Moreover, some results on sufficient conditions for the random utility representation of best-worst choices for sets of four or more alternatives are sketched.

Session: Response Times

Chair: M. Gondan

Drift Diffusion Model explains How Affective Task Content Modulates Switch Costs regarding Cue-to-Stimulus Interval in Task Switching

Maryam Sadeghi Talarposhti, Leif E. Langsdorf, Torsten Schubert

Martin Luther University Halle-Wittenberg

The impact of affective task content on task switching is an open issue. According to the affect dominance hypothesis, enhanced processing of affective task content, can lead to stronger activation of the affective task set and should result in increased switch costs for an affective compared to a neutral task. The occurrence of such asymmetric switch costs indicates that the affective task is processed as a more dominant task, which should be inhibited in order to allow for switching to the less dominant neutral task. Here, we applied a drift-diffusion model (DDM) to investigate asymmetrical switch costs during an affective task switching. Participants performed cued task switching, categorizing either gender (neutral) or facial emotions (affective) while repeating or switching tasks in two different cue-to-stimulus intervals (CTI). In line with the affective dominance hypothesis, the empirical results indicated higher switch costs for the affective compared to the neutral task. The drift rate parameter of DDM effectively captured response times and choices across all conditions and allowed to differentiate switching performance and drift rate and decision boundary as a key parameter. The affective-switch condition exhibited lowest drift rate, with higher decisional boundary reflecting larger time to switch to the affective

task. Conversely, the neutral repetition condition showed highest drift rate, and lower decisional boundary reflecting the faster switching time. Furthermore, non-decisional time representing processing time unrelated to the decision itself, increased for short cue-to-stimulus intervals. This suggests that participants needed additional processing time when emotions and task switches occurred close together, potentially reflecting the need to manage both the emotional content and the task switch itself. In our view, the reduced drift rate and increased decisional boundary in the affective-switch condition suggests a potentially increased reliance on inhibitory processes to overcome the influence of the previous neutral task, especially when dealing with emotional stimuli. This aligns with the observed switch cost asymmetry and highlights the interplay between emotional processing and task-switching mechanisms.

How to (Really) Measure the Speed of Top-down Stopping Signals

Aline Bompas

Cardiff University

People's cognitive ability for rapidly and flexibly inhibiting actions in response to sensory information is typically measured by the stop-signal reaction time (SSRT). However, I've shown recently that the SSRT largely reflects the least "cognitive" or "top-down" portion of any visuomotor reaction time: sensory and motor delays. Although widely ignored and systematically overlooked, this dependency is theoretically expected and empirically verified. I will explain why a selective stopping task is necessary to fulfil the ambitions of the SSRT, i.e. isolate the speed of top-down stopping signals. Such task interleaves signal-stop, signal-ignore and signal-absent trials. In previous work, we showed that the signal-stop and signal-ignore RT distributions depart from the signal-absent distribution at exactly the same time (Bompas et al., 2020). This "dip onset time" results from automatic inhibition and indicates the lower bound of visuo-motor delay (Bompas et al., in press). In contrast, the time when the signal-stop and signal-ignore distributions depart indicates the onset of selective stopping signals. We propose that the delay between these two landmarks provides a clean measure of stopping speed which, in contrast to SSRT, is not affected by the speed of visuomotor delays. We illustrate this using a visuo-saccadic selective stopping dataset, where the dip onset time was at 100 ms and the selective stopping started at 180 ms, suggesting it takes about 80 ms to trigger a task-specific action-withholding command in response to a stop signal. The SSRT was around 140 ms and didn't coincide with any noticeable change in behaviour.

Bompas, A., Campbell, A. E. and Sumner, P. (2020). Cognitive control and automatic interference in mind and brain: A unified model of saccadic inhibition and countermanding. Psychological Review 127(4), pp. 524-561. `https://doi.org/10.1037/rev0000181`

Bompas, A., Sumner, P. and Hedge, C. (in press). Non-decision time: the Higgs boson of decision. Psychological Review. `https://orca.cardiff.ac.uk/167265`

Interaction Contrasts for Choice Responses

Matthias Gondan

Universität Innsbruck

Townsend and Nozawa (1995, Journal of Mathematical Psychology) investigated the shape of response time distributions in two-factorial experiments for different cognitive architectures, including serial and parallel processing, with exhaustive and self-terminating stopping rules. They showed that the different architectures predict distinct shapes of the interaction contrast of the distribution functions under fairly weak assumptions, namely, selective influence of factorial manipulations on the processing times, and stochastic ordering of the processing times for different factor levels. The theory is limited to experimental tasks with ceiling accuracy, however. In this presentation, I show that with a slight extension of the stochastic dominance assumption, the original theorems can be generalized to more difficult tasks that entail non-negligible error rates (e.g., choice responses). Moreover, statistically powerful predictions can be derived for the interaction contrasts of the subdistributions of correct and wrong responses. I also apply the new method to interesting special cases such as parametric experimental variations and redundant signals tasks, and I discuss applications of the method in other areas than cognitive psychology.

Variable, sometimes absent, but never negative: Applying multilevel models of variability to the backward crosstalk effect to find theoretical constraints

Christoph Naefgen, Robert Gaschler

FernUniversität in Hagen

When performing two tasks at the same time, the congruency of the second task's features influences performance in the first task. This is called the backward crosstalk effect (BCE), a phenomenon that influences both theories of binding and of dual-task capacity limitations. The question of whether the BCE is found in all participants at all times is relevant for understanding the basis of the effect. For example, if the BCE is based on strategic choices, it can be variable, but if it is automatic and involuntary, it should never vary in whether it is present or not. Variability in observed BCE sizes was already documented and discussed when the group average effect was first reported (Hommel, 1998). Yet the theories discussed at the time did not motivate a more direct analysis of this variability, nor did the readily available statistical tools permit it.
Some statistical approaches recently applied in cognitive psychology allow such a variability-focused analysis and some more recent theoretical debates would benefit from this as well. We assessed the variability of the BCE as well as a BCE-like free-choice congruency effect by applying a Bayesian multilevel modeling approach to the data from a dual-tasking experiment. Trials consisted of a two- and a four-response task. We manipulated which task was presented first and whether the response to the four-choice task was free or forced choice.
RT data were best predicted by a model in which the BCE is zero in part of the population and drawn from a normal distribution truncated at zero (and thus always positive) in the rest of the population. Choice congruency bias data were best predicted by a model assuming this effect to be drawn from a normal distribution truncated at zero (but, in

contrast to the RT data, without the subset of the population where it is zero).
The BCE is not an inflexible and universal phenomenon that is directly linked to an inherent structural trait of human cognition. We discuss theoretical constraints implied by these results with a focus on what we can infer about the traits of the factors that influence BCE size. We suggest that future research might add further major constraints by using multi-session experiments to distinguish between-person and within-person variability. Our results show that the BCE is variable. The next step is understanding along which axes it is variable and why it varies.

Invited Keynote Lecture

Geometry of Statistical Models

Francis Tuerlinckx

University of Leuven

As quantitative psychologists, we often use statistical models to measure psychological processes and explain behavior. Geometry and measurement are at least etymologically related and geometry is deeply rooted in the measurement of quite a number of physical quantities. However, from the author's perspective, geometry has somewhat faded into the background of current psychological modeling. In this talk, I will not go into the reasons why this may the case, but I want to give an idea of how ideas from non-Euclidean geometry may be relevant for psychological modeling. We will discuss distance, curvature and volume in specific applications.

Session: Attention, Perception & Psychophysics

Chair: R. Alcalá-Quintana

Can the Queueing Model of Visual Search Account for Feature Search?

Yiqi Li

The Chinese University of Hong Kong

The queueing model of visual search (Li, Schlather, & Erdfelder, 2023) was developed for visual search in which attentive processing is necessary for the final decision, as it aims to explain the so-called attentional bottleneck, that is, the allocation of attentional resources when transiting from preattentive processing to the more resource-intensive attentive processing. Although the model's explanatory power was supported by the good model fit to empirical conjunction search and spatial configuration search data on a distributional level, whether it can account for feature search is still open. Feature search is important to understand preattentive proccessing and empirical data of feature search provide incremental information for the examination of the model assumptions. In this presentation, I will explain from a technical perspective why the adaptation of the queueing model of visual search to feature search is not a simple reduction but rather an extension. Then

I will introduce different approaches of adaptation, compare their advantages and disadvantages using simulation. Finally, fitting the adapted model to empirical data of feature search will be discussed and compared to the result of fitting to conjunction search data based on the same visual material.

On the nature of cross-modal correspondences

Jürgen Heller

University of Tübingen

There is a longstanding dispute on the nature of cross-modal correspondences. Observing transitive cross-modal matches, von Hornbostel (1931) concluded that they are absolute, irrespective of their context, and mediated by a common intensity scale shared by almost all modalities. Cohen (1934) rejected the transitivity property on empirical grounds, and suggested that cross-modal matches are relative, depending on the relative position of the judged stimuli on their respective scales. The experimental results of Marks et al. (1986) were not able to resolve the issue, as they are interpreted as falling half-way between the predictions derived from absolute and relational theories. This calls for a new theoretical perspective. Global psychophysics, in its extension to the cross-modal case (Luce, Steingrimsson, and Narens, 2010; Heller, 2021), offers a kind of hybrid approach, combining absolute cross-modal correspondences with relational judgments. Within this framework the talk will reconsider the available empirical evidence, and the implications of properties like the transitivity of cross-modal matches.

Linear Versus Logarithmic Representations of Time in Duration Discrimination

Miguel A. García-Pérez, Rocío Alcalá-Quintana

Universidad Complutense de Madrid

Psychophysical performance in duration discrimination tasks has traditionally been analyzed under a linear representation of time, that is, psychometric functions are fitted to data using chronometric time (in ms) as the independent variable. This is in contrast to Weber's law, by which discrimination thresholds are proportional to stimulus magnitude. In fellow areas of psychophysics, Weber's law is structurally incorporated into the discrimination process by considering that the independent variable for a psychometric function is the logarithm of the stimulus magnitude. Thus, in visual contrast discrimination, performance is expressed relative to log contrast; in auditory intensity discrimination, performance is expressed relative to log intensity in dB SPL. In practical terms, the importance of a logarithmic versus a linear representation of time lies in that an arbitrary (e.g., logistic) psychometric function of linear time fitted to the data will be unable to capture the actual features of data whose path is governed by log time. We present empirical evidence revealing that a logarithmic scale for time describes duration discrimination data more adequately than a linear scale. We also conducted a simulation study that fitted psychometric functions to duration discrimination data generated to obey Weber's law. The results show that fitting conventional psychometric functions (of duration in ms)

misrepresents discrimination performance and provides erroneous estimates of the difference limen, whereas fitting asymmetric psychometric functions (of log duration) captures generating performance adequately. Psychometric functions of log duration generally fitted the data much better than psychometric functions of duration in ms, although the fit turned out similar in some cases. We discuss implications for order effects, that is, for observed variations in the form and location of the psychometric function according to whether the standard stimulus is presented first or second in each trial.

A Unified Approach to Understanding Order Effects in Sensory Discrimination

Rocío Alcalá-Quintana, Miguel A. García-Pérez

Universidad Complutense de Madrid

Order effects are a deep-rooted phenomenon in psychophysical experiments. In their purest form, they occur when two stimuli differing in a single magnitude are presented successively for the observer to make a comparative judgment, and the outcome of the comparison changes with the order in which the stimuli were presented. A spatial variant has also been described when the two stimuli are displayed simultaneously at two spatial locations.

In a typical experiment, one stimulus (the standard) remains constant, while the other (the comparison) varies across trials to obtain psychometric functions that describe the probability of a given judgment as a function of the magnitude of the comparison. Studies of order effects have traditionally focused on these psychometric functions (or some derived measure of discrimination performance) as separate entities. However, psychometric functions for each presentation order are only orthogonal cross-sections of a two-dimensional surface defined by the magnitudes x.1 and x.2 of the stimuli presented first and second in each trial. Therefore, any pair of coordinates (x.1, x.2) always belongs to two orthogonal psychometric functions: one where x.1 is the standard presented first, and one where x.2 is the standard presented second. Consequently, the underlying psychometric functions for each presentation order must conform to a single mathematical expression, where the standard and comparison stimuli assume the roles of variables x.1 and x.2 as applicable. We refer to this principle as the unity constraint.

In this talk, we will discuss the scope of the unity constraint and we will present a general framework that accommodates the main models proposed to account for order effects (`https://doi.org/10.3758/s13414-020-01999-z`; `https://doi.org/10.3389/fpsyg.2017.01142`) which are compatible with the unity constraint. Additionally, we will illustrate its validity with a reanalysis of several published datasets that span different tasks and sensory modalities. Specifically, we will examine auditory discrimination of frequency modulation and visual discrimination of luminance using a same-different task, as well as visual discrimination of luminance and visual discrimination of size using a greater-less task (`https://doi.org/10.1037/a0027593`; `https://cran.r-project.org/web/packages/qdm/index.html`).

Poster Session

Stimulus Order Effects in Stimulus Discrimination Challenge Thurstonian Models of Comparative Judgment

Ruben Ellinghaus, Rolf Ulrich, Karin M. Bausenhart, Roman Liepelt

University of Hagen (FernUniversität in Hagen)

The ability to discriminate between physical magnitudes (e.g., loudness, brightness, duration) is a basic component of human judgment. Since the time of G.T. Fechner, discrimination performance has often been studied with the 2-Alternative-Forced-Choice (2AFC) paradigm, wherein which participants repeatedly compare a constant standard stimulus against a variable comparison stimulus. For the case of duration discrimination, it has been shown repeatedly that discrimination performance is better when the standard precedes rather than follows the comparison, a phenomenon which is referred to as the Type-B-Effect (TBE). This effect is not only counterintuitive but also contradicts standard psychophysical models such as Signal Detection Theory. Here, we present the results of a meta-analytic random-effects model designed to assess the TBE's generality and size. This meta-analytic regression model indicated real evidential value for the TBE and thus reveals it as a ubiquitous feature of the classic 2AFC task. Accordingly, future models of stimulus discrimination should ideally account for this effect. Although the mechanisms responsible for the occurrence of the TBE are currently not fully understood, memory updating, internal reference formation, and differential weighting of the two stimulus positions are plausible candidate mechanisms that are in line with our findings.

Enabling Large-Scale Sensitivity Analyses in Amortized Bayesian Inference with Neural Networks

Lasse Elsemüller, Hans Olischläger, Marvin Schmitt, Paul-Christian Bürkner, Ullrich Köthe, Stefan T. Radev

Heidelberg University

Sensitivity analyses are a useful tool for ensuring the robustness of computational workflows in psychology and beyond. However, they are typically forgone, as they require numerous model refits and become downright infeasible for models where even a single model fit can be computationally expensive. In this talk, we present a framework for efficiently integrating different types of sensitivity analyses into simulation-based inference with deep learning (aka amortized Bayesian inference). Our method enables efficient (i) prior and likelihood sensitivity analysis by training a single neural network for all prior and likelihood configurations of interest, (ii) data sensitivity and multiverse analysis by leveraging the fast inference of trained neural networks, and (iii) model misspecification detection by measuring the agreement within a deep ensemble of neural networks. We present experiments on representative models that underscore the effectiveness of our approach for both parameter estimation and model comparison tasks. Our results suggest that integrating sensitivity analysis into amortized Bayesian workflows is a promising step

towards reliable and robust inference.

Comparing Hypothesis Tests Using Regions of Support

Frieder Göppert, Sascha Meyen, Volker H. Franz

University of Tübingen

Hypothesis testing is one of the most widely used tools in inferential statistics. Yet, hypothesis tests - be it frequentist or Bayesian - have their respective problems and can cause sever misinterpretations. We argue that one reason for these persistent problems is the following discrepancy: While hypothesis tests are explicit on which parameter-values are theoretically contained in each hypothesis, they are usually not explicit on which parameter-values would in a practical setting lead (most likely) to which test outcome. To make these test-characteristics explicit we introduce the concept of Regions of Support (ROS) which indicate which 'true' effects likely result in which outcome of a hypothesis test given the sample size (or, more generally, precision). ROS can serve both as a check for researchers' expectations as well as a comparison of different tests. We evaluate standard Bayesian and frequentist point-null tests as well as interval (equivalence) tests on a simple, two independent samples setting. Interestingly, for interval tests our ROS analysis finds that Bayes factors suffer from an undesirable bias towards the equivalence hypothesis. We argue that other methods such as the Bayesian highest density interval (HDI) with region of practical equivalence (ROPE) or its frequentist analogue (confidence interval with ROPE) do not show this bias and might be preferable. With that, we demonstrate the diagnostic value ROS can have and hope that - due to its general applicability to any test - it will find its way into researchers' statistical toolboxes.

Lévy Versus Wiener: Assessing the Effects of Model Misspecification on Diffusion Model Parameters

Tugba Hato, Lukas Schumacher, Stefan T. Radev, Andreas Voss

Heidelberg University

Individuals display diverse decision-making behavior patterns: some prioritize accuracy, taking their time, while others tend to act swiftly out of impatience. In the case of binary decision-making, the Lévy Flight (LF) model, developed by Voss (2019), addresses these patterns by incorporating a heavy-tailed mechanism for evidence accumulation. While the Diffusion Decision Model (DDM) is commonly employed in modeling binary decision-making, we suggest that in certain scenarios, the LF model may offer a more accurate representation of the data-generation process. Hence, our aim is to investigate the parameter estimation performance of the DDM when the true data-generating mechanisms approximate the LF model. To achieve this, we conducted an extensive simulation study using simulation-based inference with neural networks as implemented in the BayesFlow framework, an approach suitable for models lacking analytical likelihood functions like the LF. Given that neural networks are susceptible to misspecification, we benchmarked our BayesFlow estimates against those of the gold-standard Stan. The comparison of parameter estimates for the standard DDM between BayesFlow and Stan revealed a close

correspondence for both DDM and LF data, thereby validating our methodology against a strong baseline. Regarding our substantive question, both BayesFlow and Stan exhibited nearly identical estimation biases when fitting the DDM to LF data: non-decision time and boundary separation parameters were underestimated, with drift rates also showing a slight underestimation effect. These findings suggest that neural networks can closely approximate the true posteriors of the DDM, but the DDM have notable biases when estimating the core DDM parameters from LF-like data.

Missing the note – modelling the interplay of temporal and sensory uncertainty in duration judgement

Theresa S. Horn, Kevin Lloyd, Peter Dayan

Max Planck Institute for Biological Cybernetics

The perception of the passage of time by humans and other animals is imprecise. This not only affects our temporal judgments, but also shapes our broader perception of the environment and our understanding of the consequences of our actions. Moreover, when temporal judgments are prompted by external sensory cues, we have to cope with additional sources of inaccuracy, such as lapses of attention. In this project, we explore the intricate relationships between uncertainty arising from temporal estimation and sensory perception. We develop a partially-observable Markov decision process model to examine how these forms of uncertainty interact, crucially incorporating the possibility of lapses of attention about which subjects might be retrospectively uncertain. We consider in detail a study that used a temporal bisection task with mice, in which they were required to categorise the delay between two auditory cues as being either long or short, relative to their past experiences of such delays. Notably, contrary to a reward rate maximising strategy, mice often delayed their responses, waiting substantial amounts of time before subsequently (correctly) reporting a delay as long. We hypothesise that this delayed responding arises from perceptual uncertainty concerning the second auditory cue. Furthermore, despite being penalised for doing so, mice repeatedly responded prematurely within the delay period prior to delivery of the second auditory stimulus, and therefore frequently failed to perform the task correctly. Premature actions followed a systematic temporal pattern, with animals typically reporting a short delay for premature responses occurring close to the first auditory cue, and a long delay at later timepoints in a trial. We hypothesise that this impulsive responding is a form of Pavlovian misbehaviour rooted in the animals' temporal expectations about action-reward pairings. Thus, we aim to model both of these apparently irrational behaviours – responses that are too early or too late – as stemming from an interplay between temporal and sensory uncertainty.

dRiftDM: An R package for Fitting Diffusion Models with Time-Dependent Parameters

Markus Janczyk, Thomas Richter, Rolf Ulrich, Valentin Koob

University of Bremen

In many areas of psychology and neuroscience, drift-diffusion models (DDMs) have become

an important framework for understanding decision processes. Models in this framework assume that response information accumulates in an incremental but noisy manner until a threshold is reached. To date, several software packages exist to fit DDMs, ranging from more classical packages such as fast-dm (Voss & Voss, 2007, BRM) or ez (Wagenmakers et al. 2007, PBR) to modern Python packages such as pyddm (Shinn et al., 2020, eLife) or PyBEAM (Murrow & Holmes, 2023, BRM). However, many of these packages are either limited to time-independent parameters or require knowledge of Python. Here we present the dRiftDM package, an R package for fitting DDMs with time-dependent parameters. The package uses a numerical approximation of the Kolmogorov forward equation to fit DDMs via maximum likelihood. The dRiftDM package is designed to be easy to use and with the typical requirements of psychological researchers in mind. For example, we provide straightforward functions for fitting and loading data sets, exploring model properties, or performing model comparison. dRiftDM can be used flexibly to implement a wide range of DDMs. In addition, it already provides pre-built models that are common in cognitive psychology. By making it easy to apply DDMs in R, dRiftDM is a valuable tool that provides researchers with an entry point to a model-driven approach to their data.

The Cognitive Processes Underlying Stereotype Threat in Gender Stereotype Domains: A Diffusion Model Analysis

Kim Keller, Mischa von Krause, Andreas Voß

Heidelberg University

Stereotypes contribute to performance differences as negative stereotypes can be perceived as threatening, especially the threat of being judged negatively or inadvertently confirming the stereotypes. When stereotypes are made salient, the stereotyped groups perform worse in many academic and non-academic domains. The primary goal of this study was to determine which cognitive processes underlie Stereotype Threat. To the best of our knowledge, no study investigated Stereotype Threat as a cognitive process using diffusion modeling and reaction times. 400 men and 400 women were asked to complete a series of trials, either a mean estimation task or an emotion recognition task. Half of the group was told that we were interested in their perceptual abilities. The other half was informed that we are testing their mathematical abilities or emotion recognition skills and that we were interested in gender differences. The manipulation in the second group revealed Stereotype Threat effect for women in mathematical tasks. A diffusion model analysis and group differences analyses revealed higher thresholds for women in the Stereotype Threat condition. The results suggest that higher motivation and more liberal decision criteria are induced by Stereotype Threat. The lack of difference in drift rates indicates no changes in working memory capacity, as many previous studies have suggested. This is an important step in explaining Stereotype Threat and preventing the impact of negative stereotypes on marginalized groups.

About cross-modal commutativity in magnitude production

Dorina Kohler, Jürgen Heller

Unversity of Tübingen

Can we compare the loudness of a tone to the brightness of a light? The answer is yes. We are intuitively capable of these cross-modal comparisons. Psychophysical researchers such as Stevens have long assumed that these cross-modal comparisons are mediated by a single scale of subjective intensity. Luce developed a psychophysical theory for physical intensity making Stevens' assumptions towards an underlying scale of perceived intensity explicit and formulating empirically testable conditions for it. They identified cross-modal commutativity as a property through which the theory can be tested. We investigated this property in a cross-modal magnitude production task between auditory and visual stimuli, concerning their loudness and brightness respectively. Participants were presented with the two stimuli and instructed to, for example "make the tone 3 times as loud [as the visual stimuli appears bright]". This was partly a replication of a paper in which the original authors concluded that cross-modal commutativity holds whereas we find inconclusive evidence in a Bayesian analysis. Additionally, in a theoretical analysis, we find evidence that role-independence of the internal references used in magnitude production is violated. In an expansion of Luce's theory, Heller concluded that cross- modal commutativity holds if and only if the internal references are role-independent, meaning they are not dependent on whether the reference pertains to the standard or the variable stimulus. This means, if role-independence of the references is violated, the assumed intensity scale can hold even if cross-modal commutativity doesn't. Evidence towards this conclusion and its implications will be discussed.

A Basic Local Independence Model for the probabilistic assessment of (un)learning items in Knowledge Space Theory

Martin Losert, Alice Maurer, Jürgen Heller

University of Tübingen

Probabilistic Knowledge Space Theory (PKST; Doignon & Falmagne, 1999) provides a set-theoretic framework for the assessment of a subject's mastery of items within a knowledge domain that can account for response errors (i.e., careless errors and lucky guesses). For longitudinal applications, skill-based PKST has been established (Anselmi et al., 2017; Stefanutti et al., 2011). In these models, skills may be gained or lost between two points of measurement, and the associated parameters for gaining and losing skills may vary between multiple groups. In the present work, we develop similar models that extend the Basic Local Independence Model with parameters for gaining (or losing) item mastery. Several methods of parameter estimation are implemented in R: Maximum likelihood (ML) parameter estimation via the expectation-maximization algorithm and, avoiding the possibility of response error inflation, both a minimum-discrepancy (MD) method and a hybrid MDML method (Heller & Wickelmaier, 2013). Results on parameter recovery

and identifiability are presented.

Modeling the discrepancy between items and knowledge states

Alice Maurer, Jürgen Heller

University of Tübingen

The most prominent probabilistic model in knowledge structure theory is the basic local independence model. One of its fundamental assumptions is the constancy of the response error probabilities across all participants. However, it seems to be implausible that a student with no knowledge in a domain guesses the correct answer of an item with the same probability as an experienced student, who masters all prerequisites of the item. Therefore it would be desirable to let an item's error probabilities depend on a person's knowledge state. Capturing some kind of discrepancy between item and state may be a starting point to achieve this. But there are various options of defining such a discrepancy measure. This poster provides an overview of existing measures and presents new ideas. The notions of inner and outer layers introduced by Doble, Matayoshi, Cosyn, Uzun, and Karami (2019) are generalized to represent more closely the length of a possible learning path from an item to a state, or vice versa.

References

Doble, C., Matayoshi, J., Cosyn, E., Uzun, H., & Karami, A. (2019). A data-based simulation study of reliability for an adaptive assessment based on knowledge space theory. International Journal of Artificial Intelligence in Education, 29 (2), 258–282.

The Limits of Measuring Metacognition

Sascha Meyen, Volker H. Franz

University of Tübingen

Searching for information in a noisy world crucially depends on knowing which things we are uncertain about. Such uncertainty assessments are a metacognitive ability and are often referred to as Type 2 ability — in contrast to the Type 1 ability of making accurate decisions. Defining appropriate measures for this Type 2 ability has gained much attention in the past 15 years. The most established measures are meta-sensitivity (meta-d') and M-ratio, which are siblings of the commonly known sensitivity (d') of Signal Detection Theory. However, these measures assume normal noise, which limits their interpretability in cases of non-normal noise distributions. Moreover, they are confounded by the Type 1 performance. That is, assessing uncertainties (Type 2) depends on how accurate predictions are in the first place (Type 1). To alleviate these shortcomings, new measures based on Classical Information Theory have been proposed recently. These measures do not need the normal noise assumption. Here, we advance the theoretical underpinning of those measures by proving that the normal noise case is only an intermediate case sandwiched between the information-theoretic worst case (binary noise) and the best case (uniform noise). In these boundary cases, the established measures are misspecified, which leads to problems with their interpretation. The information-theoretic measures, on the other hand, can successfully deal with these boundary cases. Finally, we suggest a new

information-theoretic measure that uses the boundary cases as reference points and assigns a value of 0 to the information-theoretic worst case and a value of 1 to the best case. This measure is decoupled from the Type 1 performance and can therefore be better interpreted than the existing measures. Nevertheless, problems with the interpretation of these information-theoretic measures remain. We point out such challenges and suggest future pathways to more appropriate measures of metacognition.

Representation and Assessment of Misconceptions via Knowledge Structures

Julian Mollenhauer, Jürgen Heller

University of Tübingen

Within the framework of competence-based knowledge structures, it is possible to diagnose which skills a person has mastered and which they have not yet mastered. For this purpose, a separation is made between the competence and response level of an individual and conclusions are drawn from the response behaviour via the knowledge state to the latent skills, the competence state. For this purpose, a so-called skill map is defined, which assigns the necessary skills to each item to answer it correctly. Building on a paper by Stefanutti et al. (2020), the aim is to show how, in addition to skills, associated misconceptions, i.e. a certain type of error, can also be diagnosed. The idea is to reformulate and expand this within the framework of polytomous knowledge structures according to Heller (2021). This should make it possible to arrive at a meaningful competence and knowledge structure with the help of an order on the item responses and the competence or misconception characteristics. Within the framework of a probabilistic model, a diagnosis of the latent skill or misconception can then be made, and a recommendation given as to which item should be learnt next. If there are dependency relationships between the skills or misconceptions, it should also be possible to specify the next skill to be learnt or the next misconception to be unlearnt.

How Can We Precisely Assess Collaborative and Teamwork Skills? Development and Validation of a Competence-based Knowledge Space Theory Test

Federica Morleo, Alessandra Vitanza, Pasquale Anselmi

University of Padua

To date, applications of Competence-based Knowledge Space Theory (CbKST) to test development are less common than those of Classical Test Theory and Item Response Theory, and the existing ones mainly focus on hard skill. This poster deals with the construction and validation of a test for the assessment of collaborative and teamwork skills in pre-primary and primary school teachers in the framework of CbKST. A series of hard challenges has led to specific methodological choices. First, the assessed skills need to be observable in the individuals' behaviors alone, but should involve interaction with others. Second, the test must provide a non-evaluative context so that teachers do not feel like they are being evaluated. Third, the item responses should be formatted to reduce the

likelihood of socially desirable answers, and all item responses must be independent of one another. Therefore, the test is computer-based and drawn on a situated action environment. The context is provided through narration, immersing teachers in a story where they interact with pre-programmed computer agents representing colleagues. The entire test uses a familiar interface (chat or email), where teachers respond to job situations by choosing from four alternatives or providing a short answer. For the development of the test, a collection of items (with relative response categories) was created based on the relevant skills. Each item was associated with the collaborative and teamwork skills assumed to be relevant for its resolution, forming a skill function. The test consists of a total of 20 items, which are dichotomously scored (1=correct, 0=incorrect). A sample of 140 teachers from pre-primary and primary schools in Italy was involved in the validation phase. Based on the competence-based local independence model, the validation is a complex and ongoing process, which considers multiple aspects related to: goodness-of-fit, amount of information in the data for estimating the parameters, appropriateness of the item parameter estimates for assessment ($\hat{\beta}.q + \hat{\eta}.q < 1$), and test informativeness about individuals' competence state underlying item responses. Starting from the initial skill function, created based on the skills assumed to underlie each item, various models were progressively developed, considering the outcomes listed above.

Cross-modal matching in light of Luce's global psychophysical theory

Katharina Naumann, Jürgen Heller

University of Tübingen

'Make the light as bright as the sound is loud.' This is a typical instruction in experiments dealing with the cross-modal matching of stimuli. According to Luce's (Luce, Steingrimsson, & Narens, 2010) theory of global psychophysics, in such a cross-modal task the perceived stimulus intensities are judged against respondent-generated internal reference intensities, all represented on a common psychological scale. Heller (2021) generalizes Luce's theory by distinguishing the internal references with respect to their role in the experimental setup, that is, whether they pertain to the standard or to the variable stimulus in the matching task. By testing Heller's generalization of Luce's theory of global psychophysics on cross-modal data, the present study aims at thoroughly investigating the role-sensitivity of the internal reference intensities. For achieving this, it replicates a classical experiment by Stevens and Marks (1965), who made participants adjust the brightness of a light to the perceived loudness of a noise sound and vice versa. This allows for complementing the traditional group-level analysis by evaluating the data at the individual level, and for fitting the global psychophysical model to the data in a cognitive modeling approach. We find that on the individual level, the cross-modal matching curves differ in slope, and show a regression effect as reported in the classical literature. This implies role-dependent reference intensities as suggested by Heller's model. In order to experimentally manipulate the internal references' role-(in)dependence, an alternative psychophysical method is discussed. Using an adaptive staircase procedure within the method of constant stimuli, and if instructed to choose the more intense stimulus, the subject is not aware which of the stimuli is the standard and which the variable stimulus. Under these conditions the internal references are expected to be role-independent, and the regression effect should vanish.

Heller, J. (2021). Internal references in cross-modal judgments: A global psychophysi-

cal perspective. Psychological Review, 128(3), 509–524. https://doi.org/10.1037/rev0000280

Luce, R. D., Steingrimsson, R., & Narens, L. (2010). Are psychophysical scales of intensities the same or different when stimuli vary on other dimensions? Theory with experiments varying loudness and pitch. Psychological Review, 117(4), 1247–1258. https://doi.org/10.1037/a0020174

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. Proceedings of the National Academy of Sciences of the United States of America, 54(2), 407.

Contextual effects on cross-modal matching

Annemarie Oelert, Katharina Naumann, Jürgen Heller

University of Tübingen

Cross-modal matches are two stimuli from different modalities judged equal in some relevant aspect (e.g., intensity) by an observer. For brightness and loudness, the effect of stimulus context on cross-modal matches is investigated, conceptually similar as in Marks et al. (1986). Across experimental sessions, the range of standard luminances is varied. Matched sound pressure levels are determined with a paired comparison task embedded in an adaptive staircase procedure. Instead of the expected decrease of matched sound pressure level with increasing luminance standard range (i.e. a range-dependent cross-modal normalization of perceived intensity), the data consistently show a range-dependent increase in matched sound pressure level. The global psychophysics model (Luce, 2002, 2004; Heller, 2021) fitted to the cross-modal matches describes the data adequately and can capture the stimulus context effect. Within the model, the context-dependent shift of cross-modal matches is attributed to shifts in inner reference intensities. However, as the reference intensities themselves are not uniquely identifiable, the precise nature of their shift remains unknown. Implications and further directions are discussed.

Simulation-based power analysis for psychometric functions

Maren Reiber, Florian Wickelmaier, Felix A. Wichmann

University of Tübingen

Power analysis and its importance for study planning are widely recognized. Still, it is not part of the everyday research practice in psychophysics. Consequently, it is typically unknown whether a specific experimental design is appropriate to find a relevant effect with sufficient probability. To address this issue, we ran simulation-based power analyses, considering a typical research question in psychophysics: Do thresholds differ between two experimental conditions? We specified scenarios in which two psychometric functions truly differ in their thresholds, varying design choices of the assumed data-generating model between scenarios. Here, the data-generating model is characterized by two sigmoid functions, the size of the threshold difference, and possibly an additional difference in widths. Moreover, design choices encompass the number of trials collected per function, the selection of stimulus levels as well as the task. As a result, we identified several realistic

scenarios in which power remains below the typical 80% mark, indicating the importance of power analysis. Covering a wide range of scenarios which are common in psychophysics, the resulting overview will be informative for future study planning and design evaluation.

Manual Motor Output Time: Repeatability and Correlations with White Matter

Phil Schmid, Aline Bompas, Eirini Messaritaki, Derek Jones

Cardiff University

Reaction times are a ubiquitous measure in psychological research, and while they are frequently treated as a fundamental quantity, they are at minimum composed of sensory trans-/conduction delays, decision time and motor output time (MOT). Prior modelling work suggests that decision times in a simple, speeded 2AFC task are similar across saccadic and manual response modalities, with MOT being the main contributor to the longer and more variable RT observed in the manual modality (1). Here, we present data from a subset of the WAND cohort (N = 34) who provided saccadic and manual RT to the same 2AFC task. By taking advantage of the known, fixed MOT for saccades, manual MOT could be extracted by deconvolving saccadic decision times from manual RT. Fixed, uniform, gaussian and gamma MOT distributions were evaluated, with the positively skewed gamma distribution performing the best as evaluated by goodness of fit with the empirical manual RT. Repeatability for RT and MOT parameters are presented, with means and SD for MOT distributions showing moderate repeatability. Skew, shape and delay parameters were less reliable with only scale being correlated across sessions, indicating a tradeoff between parameter values. Intraclass correlation coefficients for mean and SD for RT for both modalities were moderate. Further, based on prior work on the relationship between RT and white matter (2,3), we hypothesize that manual MOT SD should correlate with myelination in the corticospinal tract. Specifically, levels of myelin, a prime determinant of conduction velocity down axons, should demonstrate a negative correlation with MOT SD as well as MOT mean, with the latter mediated by the correlation between MOT SD and MOT mean. Preliminary results with two metrics of myelin (macromolecular proton fraction and myelin water fraction) from an ongoing experiment will be presented.

(1) Bompas, Hedge, & Sumner. 2017. Speeded saccadic and manual visuo-motor decisions: Distinct processes but same principles. Cognitive Psychology.
(2) Karahan et al. 2019. Cognitive and White-Matter Compartment Models Reveal Selective Relations between Corticospinal Tract Microstructure and Simple Reaction Time. Journal of Neuroscience.
(3) Tamnes et al. 2012. Becoming Consistent: Developmental Reductions in Intraindividual Variability in Reaction Time Are Related to White Matter Integrity. Journal of Neuroscience.

Learning in Uncertain Choice. A Reinforcement Learning Drift Diffusion Model for Probabilistic Decision-Making

Nicola Schneider, Andreas Voss

Heidelberg University

Behavioral adaptation in probabilistic environments requires learning through trial and error. While reinforcement learning (RL) models can describe the temporal development of preferences through error-driven learning, they neglect mechanistic descriptions of single-trial decision-making. On the other hand, sequential sampling models such as the drift-diffusion model (DDM) allow for mapping state preferences on single response times. We developed a joint RLDDM with a Bayesian hierarchical implementation in Stan and BayesFlow and present results from an instrumental probabilistic learning task. First, a simulation study revealed good recovery across a wide range of parameter values and only little correlation between recovered parameters. Second, we tested $N = 60$ participants and found they responded faster and more accurately with learning. The RLDDM captures this effect by modeling the drift rate through differences in Q-values and decreasing thresholds over the course of learning. Additionally, participants in the loss domain showed higher non-decision times and threshold separations than in the win domain. Lastly, for the first time, we modeled intertrial variabilities in RLDDM parameters thereby increasing model fit.

Non-stationary Reinforcement Learning Models

Lukas Schumacher, Amir Hosein Hadian Rasanan, Joerg Rieskamp, Stefan T. Radev

University of Basel

Reinforcement learning models are extensively used in various psychological research fields to explain different aspects of learning processes in decision-making. These models assume that decision agents maintain subjective value estimates for each option and learn which option is most rewarding by updating these estimates based on the prediction error — the difference between the estimated value and the actual outcome. Despite the existence of numerous model variants, two key parameters are common to most: the learning rate, which determines the weight assigned to the prediction error during value updates, and the inverse temperature parameter of the softmax choice rule, which controls the stochasticity of choices. Traditionally, these parameters are inferred from behavioral data under the assumption that they remain constant across all experimental trials. However, this stationarity assumption may be overly restrictive. Both the learning rate and the inverse temperature parameter could systematically vary throughout an experiment. For example, a decision-making agent might rely more heavily on prediction errors in the early stages of learning and increasingly trust the learned subjective values in later stages. In this study, we explore the dynamic nature of these parameters in different reinforcement learning models. To this end, we employ a recently proposed method called neural superstatistic, a modeling framework that enables us to infer plausible parameter dynamics from empirical data with minimal assumptions about their trajectories. Our findings reveal that critical reinforcement learning parameters exhibit systematic changes in various empir-

ical datasets, changes that would have remained undetected using traditional modeling approaches.

Reliability of Behavioural Interhemispheric Transfer Estimates

Heather Statham, Aline Bompas

Cardiff University

Speeded behavioural visuomotor tasks can be used to extract estimates of interhemispheric transfer time (IHTT) of neural information over the corpus callosum. The Poffenberger paradigm requires participants to give speeded responses to stimuli appearing in either visual field. If the response is made with the hand matching the side of stimuli presentation (uncrossed/ ipsilateral condition), it does not require interhemispheric transfer. When the response is given by the hand opposite to the side of stimuli presentation (crossed/ contralateral condition), it requires interhemispheric transfer. Mean or median reaction times (RTs) from these conditions are used to calculate the crossed-uncrossed difference (CUD) estimate of IHTT. However, although this measure will partly reflect the additional delay incurred by crossing hemispheres, it will also reflect any differences in the variability of RT that can result from differences in decisional processes between the crossed and uncrossed trials. Differences in visual delay would provide a more accurate estimate. Dip onset time (Bompas et al, 2024) has been proposed as a measure of the lower bound of non-decision time and should be less impacted by variability. Dip onset can be extracted using stop-signals after target onset, the presence of which produces a dip in the latency distribution since at slower RTs, participants would have enough time to process the stop-signal and inhibit their response. Theoretically, IHTT from dip onset differences between conditions should represent the time for the stop-signal to cross hemispheres and interrupt the target response before it has reached a decisional threshold from which the response can no longer be withheld. Therefore, little decisional processing time should be included within this stop-signal transfer time. The current research uses the standard Poffenberger paradigm and a stop-task variation over two sessions to compare the ability of each design to consistently provide positive IHTT estimates and sufficient test-retest reliability. The results will inform future research linking individual differences in electrophysiological and behavioural IHTT estimates.

Weber's Law for Timing and Time Perception: On the Poisson Clock

Rolf Ulrich, Karin M. Bausenhart, John Wearden

University of Tübingen

The pacemaker-counter model, a leading psychological account of timing and time perception, suggests that an internal pacemaker generates pulses according to a Poisson process, with these pulses accumulating over time. The accumulated pulses represent the elapsed duration to be timed or perceived. While the Poisson process provides a plausible mechanism for timing and time perception, it does not account for Weber's law. This discrepancy is the focus of this talk, where we propose a solution by merging Creelman's (1962) and Treisman's (1963) counting models, which can successfully account for Weber's law.

Thursday, 05 September

08:30 - 08:50			
08:50 - 09:10			
09:10 - 09:30	Decision Making I	Combining Cognitive Modeling and Eye Tracking to Investigate the Cognitive Processes Underlying Quantitative Judgments	Florian I Seitz, Rebecca Albrecht, Agnes Rosner, Jörg Rieskamp, Bettina von Helversen
09:30 - 09:50		Understanding Judgment Aggregation with Sequential Collaboration: A Computational Cognitive Modeling Approach	Maren Mayer, Daniel W. Heck
09:50 - 10:10		Coffee Break	
10:10 - 10:30	Decision Making II	Decision Heuristics and Pareto Optimality in Multi-Attribute Decision Problems	Kazuhisa Takemura
10:30 - 10:50		A Neuralistic Model of Decision Making with Plasticity, Learning, and Memory	Jerry Balakrishnan
10:50 - 11:10		Probability Distortions as Boundary Effects Under Cognitive Noise	Saurabh Bedi, Gilles de Hollander, Christian Ruff
11:10 - 11:30		Cognitive Mechanisms Underlying Outcome-Irrelevant Learning in Human Choice Behavior	Ido Ben-Artzi, Nitzan Shahar
11:30 - 12:30	Keynote	Better Accuracy for Better Science … Through Random Conclusions	Clintin P. Davis-Stober
12:30 - 14:00		Lunch Break	
14:00 - 14:20	Knowledge Structure Theory I	Toward a Unified Perspective on Assessment Models: On the Relation Between KST and CDA Models	Stefano Noventa, Jürgen Heller, Sangbeak Ye, Augustin Kelava
14:20 - 14:40		The Unidentifiability of Graded Polytomous Knowledge Structures	Andrea Spoto, Luca Stefanutti
14:40 - 15:00		A Query Procedure for Constructing Maximally Informative Tests for Skill Assessment	Pasquale Anselmi, Jürgen Heller, Luca Stefanutti, Egidio Robusto
15:00 - 15:20		Estimating the Parameters of the Simple Learning Model for Probabilistic Knowledge Structures	Florian Wickelmaier
15:20 - 15:40		Coffee Break	
15:40 - 15:50	Symposium in Memoriam of Allen Parducci	Introduction	Ehtibar N. Dzhafarov
15:50 - 16:10		Parducci Frequency	Ehtibar N. Dzhafarov
16:10 - 16:30		The „Range“ Principle in Relational Psychophysics: An Experimental Voyage of Allen Parducci	Victor Sarris
16:30 - 16:50		Contextual Effects in Salary Satisfaction	Michael H. Birnbaum, Julien Rouvere
16:50 - 17:10		The Impact of Context on Temporal Discounting	Mary Kay Stevenson
17:10 - 18:00		Business Meeting	

Session: Decision Making I

Chair: M. Mayer

Combining Cognitive Modeling and Eye Tracking to Investigate the Cognitive Processes Underlying Quantitative Judgments

Florian I. Seitz, Rebecca Albrecht, Agnes Rosner, Jörg Rieskamp, Bettina von Helversen

University of Basel

This project combines eye tracking and cognitive modeling to examine the cognitive processes underlying quantitative judgments from multiple cues. People can judge the criterion value of an object by comparing it to previously encountered, similar objects called exemplars (exemplar-similarity process) or by integrating the object's cues like a linear regression (rule process). For instance, a job applicant's suitability can be assessed based on their similarity to hired employees or based on their skill set. Often, people rely on a mixture of the exemplar-similarity and rule processes, and to investigate the role of the two processes in quantitative judgments, we made use of two methodological approaches: cognitive modeling for behavioral analyses and eye tracking for process analyses. Specifically, we tested if people who rely more on the exemplar-similarity process, as indicated by cognitive modeling, also look more at the screen locations where the exemplars were previously encoded. We conducted two eye tracking studies, in which participants judged geometric stimuli consisting of two multi-valued cues related to the criterion in either an additive (Study 1, N = 19) or multiplicative (Study 2, N = 48) way. Participants first learned the criterion value and screen location of four exemplars, each presented on a different corner of the screen. Subsequently, participants judged the criterion value of new, briefly shown test stimuli without feedback. Eye tracking captured participants' gaze proportions to the now empty exemplar locations (looking-at-nothing), and cognitive modeling within the RulEx-J framework assessed participants' reliance on the exemplar-similarity process over the rule process. The results reveal greater reliance on the exemplar-similarity process and more looking-at-nothing in the multiplicative study compared to the additive study. Focusing on the multiplicative study, participants who relied more on the exemplar-similarity process also looked more at the empty exemplar locations (Kendall's tau = 0.20, p = .05). Additional exploratory analyses showed that looking-at-nothing particularly focused on the exemplar most similar to the test stimulus and mostly occurred in the second half of the trial time after removing the test stimulus. These findings demonstrate that combining cognitive modeling and eye tracking can foster our understanding of the cognitive processes underlying quantitative judgments.

Understanding Judgment Aggregation with Sequential Collaboration: A Computational Cognitive Modeling Approach

Maren Mayer, Daniel W. Heck

Leibniz-Institut für Wissensmedien, Tübingen

Many collaborative online projects like Wikipedia use sequential collaboration to organize

the collaboration process. In sequential collaboration, contributors are presented with the latest entry, e.g., a Wikipedia article, and decide whether to adjust and maintain this entry. Recent studies examined sequential collaboration as a process for aggregating numerical judgments. They revealed an increasing accuracy of judgments along sequential chains. Notably, final estimates in a sequential chain exhibit comparable accuracy to the same number of independently aggregated judgments (i.e., wisdom of crowds). This phenomenon capitalizes at least partially on an implicit expertise-based weighting of judgments, wherein contributors modify judgments they are able to improve while maintaining judgments they are not able to improve. We combine these results and theoretical assumptions about sequential collaboration and present a computational cognitive model encompassing both sequential collaboration and individual judgments. This model not only allows to model sequential collaboration from the first initial judgment to the last contributor in a sequential chain. It also allows to model independent judgments and use their unweighted aggregate as a benchmark for performance of sequential collaboration. The model assumes an internal distribution of plausible judgments which contributors use to provide independent judgments and make decisions about the plausibility of presented judgments of previous contributors. This distribution accounts for contributors' expertise as influencing judgment accuracy. If contributors deem a presented judgment implausible, the distribution of plausible judgment is updated. It, then, also accounts for an anchoring bias introduced by the presented judgment, while also considering them as frames of reference to reduce the dispersion of provided judgments in sequential collaboration. The results of a simulation study demonstrate the capacity of our proposed computational model to replicate patterns observed in empirical studies. Furthermore, the model makes predictions about the effects of the variability in expertise, the strength of the frame-of-reference effect, and the length of sequential chains. A new empirical investigation involving extended sequential chains lends support to one of the novel predictions of the model. Overall, this computational model not only advances our comprehension of the mechanisms behind sequential collaboration, but also establishes an initial theoretical framework for further research in this domain.

Session: Decision Making II

Chair: J. Balakrishnan

Decision Heuristics and Pareto Optimality in Multi-Attribute Decision Problems

Kazuhisa Takemura

Waseda University

In the domains of behavioral decision theory and the psychology of decision making, multi-attribute decision-making problems are frequently analyzed. These problems involve decisions based on multiple attributes, often employing various heuristics (decision strategies). This study investigates these heuristics through the lens of Pareto optimality. Notably, in multi-attribute decision-making contexts, attributes are frequently assessed on an ordinal scale. Considering this, a Pareto-optimal perspective rooted in ordinal information offers significant insights. The set of Pareto-optimal solutions can serve as a valuable source of preference information to support final decision-making. In a multi-attribute decision problem, the Pareto-optimal set comprises alternatives for which no other alternative is superior in at least one attribute and superior or equal in all other attributes. This study identifies the following properties of decision heuristics for multi-attribute decision making:
(1) Alternatives selected using equal-weighted additive heuristics are proven to be Pareto solutions.
(2) Alternatives chosen by weighted-additive heuristics are Pareto solutions if all weights are positive.
(3) Alternatives selected using lexicographic decision heuristics are also Pareto solutions.
(4) In a two-stage decision-making strategy—where alternatives are initially narrowed down using a lexicographic heuristic and the final decision is made using a weighted-additive heuristic (Takemura, K., et al., 2023. Avoiding the worst decisions: A simulation and experiment. Mathematics, 11(5), p1165. DOI: 10.3390/math11051165) – the chosen alternatives are Pareto solutions.
Additionally, this study considers decision heuristics from a worst-case decision-making perspective (Takemura, K. (2021). Escaping from bad decisions: A behavioral decision theoretic approach. Perspectives in Behavioral Economics and the Economics of Behavior. London, UK: Academic Press). The discussion highlights that even simple decision heuristics can yield preferable outcomes concerning Pareto optimality, especially under worst-case scenarios. Theoretical analyses suggest that these heuristics can avoid particularly poor decisions, further demonstrating their practical utility in multi-attribute decision-making problems.

A Neuralistic Model of Decision Making with Plasticity, Learning, and Memory

Jerry Balakrishnan

California Polytechnic State University

There is solid evidence that the same area of the human brain, the midbrain Basal Ganglia, is both regulating and recording the S-R event sequences that define the outcomes of detection and recognition behaviors. The reason for this conjunction is presumably the utility of storing the relatively complex, non-verbal nature of the actual stimulus-action-result outcomes of a task in a manner homeomorphic to the neural activity that supports the behavior itself. These two neural processes must operate in vastly different time scales, with the associative memory chains being activated almost instantaneously while the perceptual/motoric sequences unfold in real time on each trial. The symbolic constructs of mentalistic (S-R) models do not lend themselves well to the analysis of the consequences of this action/representation interaction. In this paper, I present a "neuralistic" model that illustrates the basic operations that could be involved in the connections between perceptual memory and classification behavior. In addition to the mechanisms of trial-by-trial learning and the comingling of perceptual and decision-making errors, the model provides a physical explanation for the arbitrary and fickle decision-making biases that are endemic in perceptual classification tasks.

Probability distortions as boundary effects under cognitive noise

Saurabh Bedi, Gilles de Hollander, Christian Ruff

University of Zurich

Experiments consistently reveal deviations from classical economic theory, notably probability distortions in risk-related experiments. Intriguingly, similar distortions appear in perceptual judgments of relative frequency and experiments involving "complexity" without risk. We propose a comprehensive mechanistic explanation based on a domain-general model of resource-rational cognitive representation of bounded variables. Our model characterizes probability distortions observed in perception, risk, and complexity experiments as central tendency effects within the naturally bounded range of probabilities (0-1), emerging mechanistically from cognitive noise. Unlike existing Bayesian and efficient coding models, our approach doesn't assume specific prior distributions. We show that this parsimonious model captures both the distortion and variance in responses seen across experiments and directly test the causal validity of our model in an experiment on probability distortions in risky choice.
In our experiment, participants (N=63) saw lotteries with varied probabilities but fixed payoffs and reported corresponding certainty equivalents. We used a within-subject 3x2 factorial design with three probability ranges (naturally bounded from 0-1, and artificially bounded from 0-0.5 and 0.5-1) and two cognitive noise levels (complexity of presented probability information).
Our results align with all our model-predicted preregistered hypotheses. Specifically, we observed overestimation of small probabilities and underestimation of large probabilities that amplified with increased cognitive noise (p¡2e-16). Additionally, we found comparable over/underestimation above/below the artificially induced boundary at 0.5 in the high

cognitive noise condition (p¡2e-16) and an interaction effect of the artificial boundary at 0.5 and cognitive noise (p¡2e-16). Finally, we observed decreased variance in valuation close to the artificial boundary at 0.5 for high cognitive noise (p¡0.001). This confirms our central model assumptions: The characteristic inverse S-shape pattern of probability distortions causally emerges from boundary effects and cognitive noise.
Our theory predicts distortions and variance in both economic and perceptual experiments involving probabilities. However, we show with model simulations that it is applicable also to other puzzling choice phenomena, such as stake size effects and overall value effects on variability/accuracy.

Cognitive mechanisms underlying outcome-irrelevant learning in human choice behavior

Ido Ben-Artzi, Nitzan Shahar

Tel Aviv University

Reinforcement learning research shows humans use the outcomes following their actions as a learning signal. However, not all actions preceding a reward are causally related to its obtainment. Therefore, one must use prior knowledge regarding the decision environment to direct learning only for actions that are causally related to the observed outcome. Accurately assigning credit can be particularly challenging for actions that consist of multiple features from which only some are predictive of an outcome. For example, while both an apple's color and location on a tabletop are relevant for its consumption, only the color is causally related to its taste (outcome-relevant), while the location has little or no predictive information (outcome-irrelevant). Recently, studies found a replicable human tendency to assign credit to outcome-irrelevant action features. Interestingly, this tendency was found even when individuals were explicitly instructed regarding the relevancy of action features to the outcome. Here, we investigated outcome-irrelevant learning using a standard reinforcement learning four-armed bandit task. In each trial, two out of four colored cards were offered for choice, with the allocation being randomly determined by the computer. Cards led to binary rewards probabilistically according to a true expected value. Importantly, participants were explicitly made aware of the fact that only the card itself influences reward probability. Thus, they knew with high certainty that the card's location (i.e., whether it appears on the left or right of the screen) is random and therefore should not be learned. First, we replicated previous behavioral findings showing that individuals assign credit to outcome-irrelevant action features (i.e., the card's location). Next, we utilized a computational model that captured variability in the behavioral effect by parameterizing the extent to which prior instructions are used to weigh action features during choice. Finally, we used feature-specific values estimated from our model to predict participants' response times. We found that value differences between options in the outcome-irrelevant feature were associated with longer response times. Furthermore, we found that when the two features supported the same choice, reaction times were faster than when feature values were incongruent. These results indicate that humans struggle to fully inhibit learning for action features instructed as outcome-irrelevant.

Invited Keynote Lecture

Better Accuracy for Better Science ... Through Random Conclusions

Clintin P. Davis-Stober

University of Missouri

Conducting research with human subjects can be difficult because of limited sample sizes and small empirical effects. We demonstrate that this problem can yield patterns of results that are practically indistinguishable from flipping a coin to determine the direction of treatment effects. We use this idea of random conclusions to establish a baseline for interpreting effect-size estimates, in turn producing more stringent thresholds for hypothesis testing and for statistical-power calculations. An examination of recent meta-analyses in psychology, neuroscience, and medicine confirms that, even if all considered effects are real, results involving small effects are indeed indistinguishable from random conclusions.

Session: Knowledge Structure Theory

Chair: S. Noventa

Toward a Unified Perspective on Assessment Models: On the Relation Between KST and CDA Models

Stefano Noventa, Jürgen Heller, Sangbeak Ye, Augustin Kelava

University of Tübingen

In the past years, several theories for assessment have been developed within the fields of Psychometrics and Mathematical Psychology. The most notable are Item Response Theory (IRT), Cognitive Diagnostic Assessment (CDA), and Knowledge Structure Theory (KST). In spite of their common goals, these theories have been developed largely independently, focusing on slightly different aspects. Recently, Noventa, Heller, and Kelava (under review) proposed a general framework that allows to derive the models of these theories and systematize them within a general taxonomy by means of two primitives (structure and process) and two operations (factorization and reparametrization). In this contribution, the introduced framework is used to derive both KST and CDA models based on dichotomous latent variables, thus achieving a two-fold result: On the one hand, it settles the relation between the frameworks; On the other hand, it provides a simultaneous generalization of both frameworks, thus providing the foundations for the analysis of more general models and situations.

The Unidentifiability of Graded Polytomous Knowledge Structures

Andrea Spoto, Luca Stefanutti

University of Padova

The generalization of the Knowledge Structure Theory (KST) to polytomous items has paved the way for studying whether and how the deterministic and probabilistic characteristics and problems of KST apply to the polytomous case. Among these issues, the present contribution aims to investigate whether and how the problem of unidentifiability of the Basic Local Independence Model (BLIM) also applies to its polytomous extension, the Polytomous Local Independence Model (PoLIM). It has already been shown that forward and backward gradedness of the knowledge structure — a characteristic sufficient for the unidentifiability of the BLIM — also exists in the polytomous case and leads to the unidentifiability of the PoLIM. In this talk, we (i) generalize the definition of the transformations that link two equivalent PoLIM instances; (ii) define the interval in which such transformations are within the parameter space of the PoLIM; and (iii) prove that the unidentifiability of the BLIM due to forward and backward gradedness is a special case of the unidentifiability due to the gradedness of the PoLIM.

A Query Procedure for Constructing Maximally Informative Tests for Skill Assessment

Pasquale Anselmi, Jürgen Heller, Luca Stefanutti, Egidio Robusto

University of Padua

Competence-based knowledge space theory provides the theoretical framework for assessing the latent set of skills an individual has available (referred to as the "competence state") from the observed responses to test items. A recent approach to test development proposed within this framework exploits concepts originally introduced in rough set theory to construct tests that are as informative as possible about individuals' competence states (i.e., adding any item does not make the tests more informative) and minimal (i.e., no item can be removed without making the tests less informative). In the tests under consideration, each item is associated with a set of skills, and mastering all of them is necessary to solve the item. Existing procedures for test construction assume that the particular sets of skills for which an expert can provide items are known in advance. Such an assumption is not realistic in many practical applications. The talk presents a query procedure for test construction that is based on asking an expert whether or not they can provide items that require specific sets of skills to be solved. The expert is asked a limited number of questions, with each subsequent question being specifically chosen based on the answers to the previous ones. The query terminates when asking additional questions would not result in a more informative test. The resulting tests are maximally informative given the expert's capacity. Exemplary applications of the procedure are presented and discussed.

Estimating the Parameters of the Simple Learning Model for Probabilistic Knowledge Structures

Florian Wickelmaier

University of Tübingen

Knowledge structure theory seeks to provide procedures for the effective diagnosis of the knowledge state of a student in a certain domain (such as algebra, physics, or statistics; Doignon & Falmagne, 1999). The most popular probabilistic model is the basic local independence model (BLIM). One of its challenges is that its number of knowledge state parameters potentially grows very large. The simple learning model (SLM) restricts the number of state parameters by assuming a learning process where a student may propagate step by step from a state of ignorance to full mastery. The SLM includes problem-specific solvability parameters and uses them to compute the distribution of knowledge states. This presentation introduces three estimators for the SLM parameters: maximum likelihood (ML) estimators via the expectation-maximization algorithm, minimum discrepancy (MD) estimators, and MDML estimators. The estimators are implemented in the pks package in R, and their properties are illustrated in application examples and simulations. Doignon, J.-P., & Falmagne, J.-C. (1999). Knowledge Spaces (Springer, Berlin).

Symposium in Memoriam of Allen Parducci

Parducci Frequency

Ehtibar N. Dzhafarov

Purdue University

Allen Parducci was a brilliant psychophysicist and a practical philosopher of impressive depth. He was not a mathematician, however, and the math he used in his work was of elementary variety. Nevertheless, he was engaging into most delightful conversations about math-intensive problems, such as theory of measurement and nature of physical laws, and his wisdom and intuition often led him to surprisingly penetrating observations, offered in his hallmark soft-spoken and unassuming manner. Thinking of all such exchanges throughout my 40 years of knowing Allen, I chose for this presentation a very simple mathematical topic, primarily because I have a written record of it, and because it is easier to present than philosophical issues. In 2013, Allen asked me of the way of assigning cumulative frequency (CF) to rank-ordered data, subject to the following constraints: (1) the first observation has the CF of 0; (2) the last observation has the CF of 1; and (3) the mean CF in the distribution is 1/2. He had long since devised his own formula, but he was concerned he could not find it in the literature. His formula was $CF(x) = (R(x)-1)/(N-1)$, where $R(x)$ is the ordinal position of the observation x in the list of N sorted observations. As we quickly established in our discussion, none of the known to us standard ways of defining a distribution function has the desired properties. Thus, in the English-language literature, $CF(x) = \#[X \leq x]/N$ or $(\#[X < x]+\#[X = x]/2)/N$, where $\#$ stands for the "number of" and X represents the random variable we are dealing with. At the same time, an absolutely continuous distribution on an interval $[a, b]$ does satisfy all three Parducci requirements. The Parducci frequency therefore is a nice finite-set analogue

of absolutely continuous distributions. Why is then it is not used universally? The reason is that $R(x)$ is not a function of x: if there is a repeated observation $x, x, x \ldots$ the rank of each subsequent x increases by 1. This makes it problematic to a mathematician but is completely irrelevant to an experimental psychologist. Indeed, why should this be a problem if there is a rigorously defined algorithm by which $R(x)$ is computed? Moreover, the sorting procedure transforms a set of data into an indexed set, so that $x, x, x \ldots$ is, in fact, $(x, k), (x, k+1), (x, k+2), \ldots$ and $R(x, m) = m$ is a well-defined function.

The „Range“ Principle in Relational Psychophysics: An Experimental Voyage of Allen Parducci

Victor Sarris

Goethe University

This paper deals with the range principle of Allen Parducci's range-frequency (RF) model, which is based on a longstanding experimentation in relational psychophysics (Parducci, 1963, 1965). The success of the RF model rests on an interplay of fruitful mathematical model building, extensive empirical research and creative theorizing in psychophysics and cognition (Sarris, 2006; Townsend, 2008). Two illustrations are provided which stem from Allen Parducci's exchanges with this author: The first example concerns the joint experimental work on an extension of Parducci's range principle and Sarris' remote-anchoring paradigm (Sarris & Parducci, 1978); the second illustration is taken from Parducci's article on „perceptual and judgmental relativity“ as published in the book co-edited by Parducci and this author (Sarris & Parducci, 1984, reissued in 2024). The concluding remarks are provided in the light Allen Parducci's (1925-2023) prolific life and work.

References:
Sarris, V., & Parducci, A. (1978). Multiple anchoring of category rating scales. Perception & Psycho-physics, 24, 35-39.
Parducci, A. (1984). Perceptual and judgmental relativity. In V. Sarris, & A. Parducci (Eds.), Perspectives in psychological experimentation: toward the year 2000 (pp. 135-149). Hillsdale, NJ: Erlbaum (Reissued – London: Routledge, 2024).
Townsend, J. T. (2008). Mathematical psychology: prospects for the 21st century. Journal of Mathematical Psychology, 52, 259-271.

Contextual Effects in Salary Satisfaction

Michael H. Birnbaum, Julien Rouvere

California State University, Fullerton

We present a series of studies of judgments of satisfaction with salary, manipulating the distribution of salaries of others doing the same work. The experiments were designed to compare six theories of contextual effects in judgment, including adaptation level theory, correlation-regression theory, inferred distribution theory, decision by sampling, ensemble theory, and range-frequency theory. Manipulations of the frequency distribution using cubic density functions produces a double crossover of curves relating judgments to salaries; this double crossover violates implications of four of the theories but remains consistent

with decision by sampling and range-frequency theories. Inferred distribution theory assumes that rank is inferred from the mean and endpoints, so it fails to describe the double crossover. Manipulation of the endpoints produces changes in the heights and slopes of the curves, which are not explained by decision by sampling and are partially inconsistent with ensemble theory. Ensemble theory implies no effect of the rank of a salary and assumes that endpoints only affect judgments of salaries on the same side of the mean, contrary to the results. Range-frequency theory implies that ratings of stimuli holding the same ranks in two contexts with differing endpoints should be linearly related, and the data appeared consistent with this implication. Range-frequency theory is the only theory that gives a consistent account of all of the results. Range-frequency theory can be extended in order to estimate the effective context, which appears to differ systematically between people according to their full-time incomes.

The Impact of Context on Temporal Discounting

Mary Kay Stevenson

California State University East Bay

Temporal discounting describes the impact of delay on gains and losses. Context effects (Parducci, 1965, Birnbaum & Rouvere, 2023) describe the impact of the stimulus range on the judged interpretation of stimuli. The current study assesses the impact of several contextual manipulations on the temporal discounting of gains and losses. The delay range affects the impact of individual delay weights. The same delays had different discounting weights depending on whether they were the shortest or longest delays in the range of options. Gains and losses have different discounting rates whether they are isolated or judged as a combined outcome. The subjective values of gains and losses are influenced by the outcome format. These results have implication for short term gains and long term consequences that are ignored.

Friday, 06 September

08:30 - 08:50	Modeling I	An Empirical Test of the Two-High-Threshold Contrast Model cancelled	Marie Jakob Constantin G. Meyer-Grant
08:50 - 09:10		Extending Multinomial Models of the Illusory Truth Effect with Response Times: A Comparison of RT-MPT Models	Raphael Hartmann, Constantin G. Meyer-Grant, Lars Kulbe, Daniel W. Heck
09:10 - 09:30		Empirical Versus Normative Weights of Sequentially Sampled Advice	Tobias R. Rebholz, Mandy Hütter, Andreas Voss
09:30 - 09:50		Disentangling Conditional Dependencies	Nicole Cruz, Michael D. Lee
09:50 - 10:10		Coffee Break	
10:10 - 10:30	Modelling II & KST II	Bayesian and Heuristic Approach to Assess Human Decision Making and Confidence: A Computational Replication Study	Cem Tabakci, Sebastian Hellmann, Michael Zehetleitner, Manuel Rausch
10:30 - 10:50		Using Interpretable Machine Learning for Differential Item Functioning Detection in Psychometric Tests	Elisabeth Barbara Kraus, Johannes Wild, Sven Hilbert
10:50 - 11:10		Deep Sequential Sampling Model: A Data-Driven Cognitive Modeling Approach	Amir Hosein Hadian Rasanan, Florian Seitz, Nathan J. Evans, Joerg Rieskamp
11:10 - 11:30		From Knowledge Structures to Preference Structures: Extending the basic local independence model to forced-choice pair comparison tasks	Debora de Chiusole, Luca Stefanutti, Andrea Brancaccio
11:30 - 12:30	Keynote	A United Discipline: Hierarchical Modeling of Individual Differences in Cognitive Tasks	Julia Haaf
12:30 - 13:00		Closing Session	

Session: Modeling I

Chair: T.R. Rebholz

An Empirical Test of the Two-High-Threshold Contrast Model

Marie Jakob, Constantin G. Meyer-Grant

University of Freiburg

A longstanding debate in memory research revolves around the question whether recognition memory judgements are best conceptualized as resulting from the direct comparison of a latent memory signal with a response criterion or through a mediation of memory signals by a small number of latent states. These perspectives have commonly been represented by signal detection theory (SDT) and the two-high-threshold model (2HTM), respectively. Kellen and Klauer (2014) showed that common SDT models and the 2HTM make conflicting predictions in a ranking paradigm and implemented a critical test on that basis; their results were in line with the predictions made by the SDT models and contradicted the 2HTM. However, this conclusion was recently called into question by Malejka and colleagues (2022) who proposed that recognition decisions involving multiple stimuli are based on a contrast mechanism. They argued that if the detection probability for any stimulus in a given set is determined by comparing the memory strength of the stimulus in question with the memory strength of the other stimuli in the set, the 2HTM is able to account for the results of Kellen and Klauer (2014). In order to assess whether a 2HTM that incorporates such a contrast mechanisms is empirically adequate, we directly tested one of its key predictions using a ranking paradigm and found evidence against the model. By contrast, our results align well with the predictions made by SDT. We discuss implications of our results for models of single- and multiple-item recognition memory.

Extending Multinomial Models of the Illusory Truth Effect with Response Times: A Comparison of RT-MPT Models

Raphael Hartmann, Constantin G. Meyer-Grant, Lars Kulbe, Daniel W. Heck

Albert-Ludwigs-Universität Freiburg

The illusory truth effect (ITE) refers to the finding that statements, which have been repeatedly encountered by participants, are judged "true" more often than statements which have not been encountered (even if the statements are actually false). Based on a model comparison between two multinomial processing tree (MPT) models, Fazio et al. (2015) claimed that knowledge does not protect against the ITE. Both of their MPT models assume three processes: familiarity, knowledge, and guessing. The two models differ in the assumption of whether familiarity or knowledge is the dominant process (i.e., whether familiarity can override the response produced by the knowledge-process or vice versa). In their model comparison, Fazio et al. (2015) found that the MPT model with the fluency process being dominant fits the data better, hence their concluding claim. We aim at investigating the mechanisms behind the ITE further by integrating response times into the MPT models (i.e., using RT-MPT models), which allows to make statements about the order of latent processes, and by using a slightly different design; We use three instead

of two knowledge conditions (i.e., widely known, moderately known, and widely unknown) and two conditions for the false statements (plausible and implausible). In addition, we consider additional (RT-)MPT models and compare all models using different information criteria (LOOIC, WAIC, and DIC) to find the model that describes the data best while considering the flexibility of each model. With this project, we want to (1) check whether we can reproduce the finding by Fazio et al. (2015) while accounting for response times, (2) provide potentially alternative explanations for it, and (3) get a deeper insight into the cognitive mechanisms and processes involved in the ITE.

Empirical Versus Normative Weights of Sequentially Sampled Advice

Tobias R. Rebholz, Mandy Hütter, Andreas Voss

University of Tübingen

In the traditional advice taking paradigm, the judge-advisor system, participants are presented with only a single piece of external evidence as advice after they have formed their own initial beliefs. In sampling extensions, by contrast, participants can sequentially sample multiple pieces of advice before making their final judgment. Normative strategies for integrating sequentially sampled information from multiple external sources can be derived based on different accounts of Bayesian belief updating. In particular, the traditional modeling strategy can be shown to implement sampling invariant simple cumulative averaging, that is, the assumption that participants weight all pieces of external evidence equally in their End-of-Sequence compromising. In contrast, we also develop a Step-by-Step updating model that implements a sampling non-invariant two-stage processing that sequentially alternates between (re-)assessing the perceived validity of others' judgments in stage one and updating participants' own beliefs about the truth in stage two. To contribute to the understanding of active advice seeking, these two types of normative weights are compared with separate empirical weights estimated based on corresponding multivariate mixed-effects regression modeling. In a reanalysis of empirical data, we find evidence for significant order effects, and thus against the traditional equal-weighting approximation, for participants compromising between their own initial beliefs and multiple pieces of advice sequentially sampled from others. Indeed, there is a strong positive correlation between the corresponding empirical and normative Step-by-Step weights of sequentially sampled advice, but the correlation is stronger for the End-of-Sequence weights. Moreover, similar to the traditional single advice taking paradigm, there is evidence for egocentrism in the sense of serial underweighting of external evidence. However, there are also large inter- and intra-individual differences in strategy selection for sequential advice taking. Indeed, some participants behave almost perfectly in line with the predictions of Bayesian belief updating. The normative and empirical models of sequential advice seeking can also be applied to investigate individual differences in participants' sampling decisions. Our findings constitute initial evidence for the adaptive utilization of multiple, sequentially sampled external opinions.

Disentangling Conditional Dependencies

Nicole Cruz, Michael D. Lee

University of Potsdam

People draw on event co-occurrences as a foundation for causal and scientific inference, but in which ways can events co-occur, and can dependencies between events be combined to form more complex dependencies? Statistically one can express a dependency between events A and C as P(C given A) != P(C), or P(A given C) != P(A). But this relation can be specified further, yielding cases in which the above two terms are not equivalent. In the psychology of reasoning, biconditional relationships have often been thought to occur when people add to the probability of a conditional P(C given A) the converse, P(A given C), or inverse, P(not-C given not-A), or both, with the effects of these additions largely treated as equivalent. But from a coherentist perspective it can make a strong difference whether the converse or the inverse is added, and in what way. In particular, the addition can occur by forming the conjunction of the two conditionals, or by merely constraining their probabilities to be equal. Here we outline four distinct ways of defining biconditional relationships, and illustrate their differences by how they constrain the conclusion probabilities of six inference types. We present a Bayesian mixture model with which the biconditionals can be dissociated empirically from one another, and discuss implications for the interpretation of empirical findings in the field.

Session: Modeling II & Knowledge Structure Theory II

Chair: E.B. Kraus

Bayesian and Heuristic Approach to Assess Human Decision Making and Confidence: A Computational Replication Study

Cem Tabakci, Sebastian Hellmann, Michael Zehetleitner, Manuel Rausch

Catholic University of Eichstätt–Ingolstadt

Due to the ongoing replication crisis in psychology, it has been suggested that psychologists should make more widespread use of formal cognitive modelling. However, the many decisions researchers need to make during cognitive modelling raise doubts about whether results based on cognitive modelling will be replicated. Here, we present a replication attempt of Adler and Ma's (PLoS Comp Biol, 14(11), e1006572) finding that heuristic models outperform Bayesian models in an orientation discrimination task with simultaneous confidence judgments, suggesting that human perceptual decisions are not Bayes-optimal. After processing the authors' original data and reprogramming the modelling analysis from the ground up, our analysis replicated Adler and Ma's central finding that heuristic models better fit the data than the Bayesian models. However, our versions of the modelling analysis produced worse model fits than the results reported by Adler and Ma. In general, it doesn't seem convenient to replicate computational cognitive models without comprehensive details about the involved computations and the underlying computer

code.

Using Interpretable Machine Learning for Differential Item Functioning Detection in Psychometric Tests

Elisabeth Barbara Kraus, Johannes Wild, Sven Hilbert

LMU Munich

This contribution explores a novel methodology combining psychometrics and machine learning to assess test fairness and identify differential item functioning (DIF). DIF occurs when different demographic groups (e.g., based on race, gender, etc.) respond differently to specific test items despite having the same underlying ability level. Traditional methods for detecting DIF, such as the Mantel–Haenszel statistics and logistic regression, have been effective but do not readily account for complex relationships between demographic variables and item responses nor to multi-dimensional factor structures. We therefore propose a machine learning approach that leverages interpretable models applied to random forest models to uncover complex, non-linear relationships between test items and demographic variables. This method evaluates the importance of individual test items and latent ability scores by comparing their predictive power against a random baseline variable when predicting demographic characteristics. To validate our method, we conducted simulation studies under various conditions, such as such as linear and complex impact and unfairness, as well as varying number of factors, unfair items, and varying test length. Our findings suggest that the machine learning approach is as reliable as traditional methods for detecting unfair items and offers enhanced generalizability to multidimensional scales. We demonstrate the practical application of our method using an elementary school reading comprehension test, where random forests were employed to predict students' migration backgrounds from their test responses. The analysis identified one item as unfair according to all proposed decision criteria, with content analysis providing plausible explanations for this finding.
This study highlights the potential of machine learning to enhance the detection of DIF by being sensitive to more complex data structures and relationships, ultimately contributing to the development of fairer psychometric assessments. For further details and access to the analysis code, you can visit the project page on OSF: `https://osf.io/s57rw/?view.only=0440c630cbd5489f85232a30a37f1b8a`.

Deep Sequential Sampling Model: A Data-Driven Cognitive Modeling Approach

Amir Hosein Hadian Rasanan, Florian I. Seitz, Nathan J. Evans, Joerg Rieskamp

University of Basel

Understanding the mechanisms underlying decision making is crucial for many fields of science like psychology, economics, neuroscience, and cognitive science. In the last decades, scientists have extensively used cognitive modeling methods to formalize verbal theories of decision making and get deeper insights into the underlying cognitive mechanisms. Sequential sampling models (SSM) are among the most popular cognitive decision-making models. These models assume that the decision maker integrates noisy evidence for each

available option at a constant rate (i.e., drift rate) until the relative or absolute amount of evidence to the benefit of one option reaches a threshold. For explaining human behavior, cognitive scientists usually propose a specific functional form (i.e., mapping function) that explains how stimulus features (e.g., attribute values of the presented items) and cognitive constructs like attention modulate different parameters of the decision model (e.g., the drift rate). Although specifying the functional forms turns a verbal theory into a testable model, it also harbors the risk of misspecification, in which important aspects of human behavior are not taken into account. This project aims to provide a data-driven modeling approach called the deep sequential sampling model (DeepSSM), which utilizes a deep neural network that learns the mapping function between measurable features and latent parameters of the SSM from behavioral data. One benefit of this approach is that the cognitive scientist does not need to have a prior assumption about the functional form of the mapping function, and the model can learn it from the data. Besides, DeepSSM can be adapted to new observations. To test the proposed DeepSSM, we have fitted the model on data from five different choice experiments with different setups (perceptual versus preferential domain, binary versus multi-alternative choice, and uni-attribute versus multi-attribute stimuli). In each experiment, the neural network took participants' gaze proportions and their subjective valuations of the different alternatives as input and predicted the drift rate, which then served as input for a race diffusion model. The proposed model can predict the observed patterns (e.g., the difficulty effect on response time and choice or the gaze effect on choice) in the empirical data without prespecifying any functional form to predict those patterns. Also, the out-of-sample predictions of the model after training on only 60% of the data showed that the model has a strong generalization power for the unobserved data. Moreover, since DeepSSM is a combination of a deep neural network and an SSM, it inherits the interpretability of SSMs.

From Knowledge Structures to Preference Structures: Extending the Basic Local Independence Model to Forced-Choice Pair Comparison Tasks

Debora de Chiusole, Luca Stefanutti, Andrea Brancaccio

University of Padua

The basic local independence model (BLIM) is a probabilistic model originally developed for empirically validating knowledge structures. To date, nearly all applications of BLIM to real data have focused on knowledge assessment. This research extends the BLIM to forced-choice pair comparison tasks. In these tasks, given a set A of options, an individual must establish a preference between any two options a and b in A, considering all possible pairs (a, b). Building on the "true and error" (TE) models proposed by Birnbaum and colleagues, it is assumed that each individual has a "true preference" represented by an asymmetric and connected relation within set A. This true preference is called a *preference state*, and the collection of all preference states is termed the *preference structure*. In practical scenarios, individuals may commit errors. Thus, for every possible pair (a, b), an "error rate" e_{ab} is defined as the probability of observing "b preferred over a" given that the true preference is "a preferred over b". This study introduces the correspondence between the primitives and assumptions underlying both the BLIM and the TE models, leading to an adaptation of the BLIM for forced-choice pair comparison tasks. A key

aspect of the TE models involves the *replication* of the same choice problems, where the same choice problem is administered to subjects multiple times. The non-coherence of subjects' choices across these replications forms the basis for computing the error term for each comparison. Results from a simulation study clarify when replications are necessary to obtain reliable estimates of the model parameters and when they are not. Additionally, new findings regarding the identifiability of the BLIM suggest that some identifiability issues common in knowledge assessment studies do not occur in the study of preferences.

Invited Keynote Lecture

A United Discipline: Hierarchical Modeling of Individual Differences in Cognitive Tasks

Julia Haaf

University of Potsdam

Hierarchical modeling is an incredibly useful and widely used approach for repeated measures designs where many observations are nested within participants. In such designs, hierarchical models can optimally account for the structure of the data and thus reduce estimation bias. However, an obvious advantage of hierarchical modeling is rarely exploited: By explicitly separating inter-individual variability from intra-individual variability, hierarchical modeling can improve our understanding of individual differences. In this talk, I will show how hierarchical modeling serves as a psychometric tool for investigating individual differences in cognitive phenomena. First, using attentional control as an example, I will discuss how limited our understanding of individual differences in this domain is and how hierarchical modeling might help to improve the current state of affairs. Second, I will discuss how the nature of (some) cognitive tasks may limit what we can learn about individual differences, even with hierarchical modeling.

List of Authors

Program Schedule

	Wednesday	Thursday	Friday
08:30 - 09:10	Opening & Measurement		
09:10 - 09:50		Decision Making I	Modeling I
09:50 - 10:10	Coffee Break	Coffee Break	Coffee Break
10:10 - 11:30	Response Times	Decision Making II	Modeling II & KST II
11:30 - 12:30	Keynote Francis Tuerlinckx	Keynote Clintin Davis-Stober	Keynote Julia Haaf
12:30 - 14:00	Lunch Break	Lunch Break	Closing Session
14:00 - 15:20	Attention, Perception & Psychophysics	Knowledge Structure Theory I	
15:20 - 15:40	Coffee Break	Coffee Break	
15:40 - 17:00	Poster Session	Symposium in Memoriam of Allen Parducci	
17:00 - 18:00	Young Scientists Networking Event	Business Meeting	