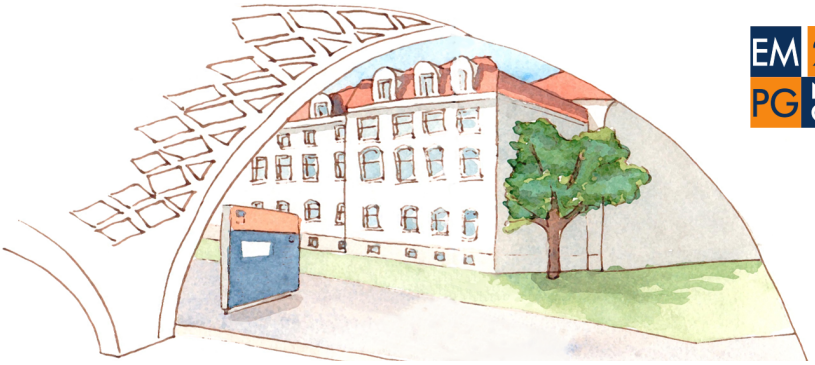
European Mathematical Psychology Group 2026

Monday, 7.9.2026 – Wednesday, 9.9.2026

University of Innsbruck



Book of Abstracts

Contents

Poster / 18

Analyzing Binary Judgments: Comparing Two-Step ANOVA and Generalized Linear Mixed Models from the Perspective of Signal Detection Theory

Author: Semih C. Aktepe[1]

Coauthor: Daniel W. Heck [1]

[1] *Philipps-Universität Marburg*

Binary judgments are widely used in psychological research to investigate cognitive processes and decision making. Statistical frameworks can be divided into aggregation-based approaches like analysis of variance (ANOVA) and trial-level analysis approaches like generalized linear mixed models (GLMM). The latter can be framed as signal detection theory (SDT) models, providing more insight into decision-making processes in terms of individuals'bias and discrimination ability. For substantive researchers, it may be unclear how these approaches differ in terms of model specification, robustness, and results. Given that a systematic comparison using empirical datasets remains lacking, it is difficult to judge the practical relevance of choosing a specific approach. To address this gap, we compared ANOVAs and GLMMs along with their SDT interpretations using 20 openly available datasets featuring the illusory truth effect, a robust phenomenon frequently examined with binary judgments. We systematically compared seven GLMM versions of increasing random-effects complexity against two ANOVA-based criteria—the by-subjects F1, the by-items F2, and their combination via the min-F'statistic—and a two-step SDT approach. We also conducted a Monte Carlo simulation examining statistical power, Type I error rates, and parameter recovery across all approaches under varying sample and effect sizes. The results showed that GLMMs with a moderately complex random-effects structure produced more stable and conservative effect size estimates than the other methods. Moreover, GLMMs can resolve assumption violations of ANOVA and enable a direct interpretation in terms of SDT while being less sensitive to missing data. Overall, GLMMs are a theoretically sound and practically robust method and thus superior for analyzing binary judgments in social and cognitive psychology.

Topic:

Statistical methods

Talks / 26

Proper units for pro-environmental behavior: a review and feasibility study of private-sphere behaviors

Author: Irene Alfarone[1]

Coauthor: Matthias Gondan [1]

[1] *University of Innsbruck*

One major focus of environmental psychology is on Pro-Environmental Behavior (PEB); however, there is little consensus on how it should be measured. In this study, we reviewed articles published in the *Journal of Environmental Psychology* in 2023 and examined how private sphere PEB was operationalized. We then translated the reported behaviors into a common metric, tCO_2-eq, and conducted a meta-analysis of their potential CO_2 savings. The results show substantial heterogeneity in measurement practices and a strong predominance of low impact behaviors. Existing self-report instruments often combine behaviors that differ widely in frequency, difficulty, and environmental impact, despite being used as operationalizations of the same construct. This raises concerns about the interpretation of PEB scores and the extent to which they can meaningfully contribute to cumulative science, reflecting a broader measurement problem that is common across many areas of psychology.

Topic:

Measurement and scaling

Poster / 6

Trial-aligned joint EEG-behaviour evidence accumulation modelling enables cross-modal prediction in time-pressured pedestrian road-crossing decisions

Authors: Jamal Amani Rad[1]; Gustav Markkula[2]; Thomas O. Hancock[2]; Stephane Hess[2]

[1] *University of Leeds, UK*

[2] *University of Leeds*

Many neurocognitive modelling approaches still connect EEG and behaviour only after separate analyses, which weakens trial-level mechanistic claims and limits prediction. We present a trial-aligned joint model that treats binary crossing choice, response time (RT), and single-trial centro-parietal positivity (CPP) as co-generated by shared evidence accumulation dynamics. The behavioural core is an accumulation-to-bound process producing first-passage decisions and RTs under time pressure (operationalised via time-to-arrival, TTA). CPP is modelled as a noisy readout of a decision-relevant latent evidence summary in a fixed pre-response window, with participant-specific coupling parameters. Because the full multimodal likelihood is intractable, we use amortised simulation-based Bayesian inference to obtain hierarchical posteriors over cognitive and neural parameters.

We apply the model to a controlled pedestrian road-crossing experiment ($N = 16$) with repeated TTA manipulations (2.5–4.0 s), yielding trial-wise choices, RTs, and CPP features. Posterior predictive checks show that a single parameterisation reproduces the full RT distribution for cross-before and cross-after trials across TTA levels, while simultaneously capturing condition-dependent shifts in CPP. Trial-wise coupling is supported by a robust positive association between CPP and the latent evidence summary ($r = 0.51$). The joint formulation enables genuinely cross-modal prediction: adding an early CPP window improves EEG $\rightarrow$ choice discrimination beyond a TTA-only baseline (AUC 0.92 vs 0.66), while predicting CPP from behaviour yields calibrated predictive intervals at the single-trial level. Crucially, performance drops under an EEG-behaviour trial-shuffling control (AUC $\approx$ 0.70), providing a direct diagnostic that the model exploits within-trial alignment rather than only condition-level covariation.

Beyond this proof-of-concept, the framework supports principled model comparison of alternative neural commitments (e.g., CPP linked to drift, boundary proximity, or urgency) within one joint posterior predictive space. Overall, we provide a bridge from mathematical psychology evidence accumulation theory to a more naturalistic, safety-critical decision, while retaining identifiability checks and uncertainty-aware prediction.

Topic:

Mathematical models of psychological processes

Talks / 9

Using a mathematical representation of brain processes to explain decision making: adapting Friston's free energy principle for travel behaviour modelling

Authors: Jamal Amani Rad[1]; Thomas O. Hancock[2]; Stephane Hess[3]

[1] *University of Leeds, UK*

[2] *University of Leeds*

[3] *T.O.Hancock@leeds.ac.uk*

This paper adapts the Free Energy Principle (FEP) to route choice under uncertainty and examines whether an active inference account can explain travel learning better than conventional reinforcement learning (RL). We develop a route choice model in which travellers update beliefs about hidden traffic conditions and select routes by minimising expected free energy, thereby jointly capturing preferences, uncertainty, and exploration. The model is estimated in a hierarchical Bayesian framework and applied to two complementary experimental datasets: an incentive-compatible driving simulator study and a laboratory route choice study with information manipulations. Across both datasets, the FEP model provides a better account of observed behaviour than RL, especially for participants showing exploratory or belief-adaptive responses. The results suggest that active inference offers a promising neurocomputational foundation for richer models of adaptive travel behaviour.

Topic:

Models of cognition and learning

Poster / **7**

EZ-SDM: Closed-form estimation for spherical diffusion models in 3D

Author: Jamal Amani Rad[1]

[1] *University of Leeds, UK*

Hyperspherical diffusion models extend evidence-accumulation theory to continuous, multidimensional report spaces, where decisions correspond to first passage to a spherical boundary and responses are given by the hitting direction on the sphere (Smith & Corbett, 2019). Despite their theoretical appeal, parameter estimation often requires computationally intensive likelihood evaluations (e.g., integral-equation solvers; Hadian Rasanan et al., 2025), limiting routine use in large datasets and hierarchical workflows. We introduce EZ-SDM, a fast, moment-based estimator for the three-dimensional spherical diffusion model with a fixed threshold, inspired by EZ-CDM for circular diffusion (Qarehdaghi & Rad, 2024). The core observation is that the joint distribution of decision time and response direction factorizes: response directions follow a von Mises Fisher distribution on the unit sphere, with concentration determined by the product of boundary radius and drift magnitude (scaled by diffusion), while decision-time moments depend on the same concentration through closed-form expressions. Leveraging this structure, EZ-SDM yields explicit estimates of drift magnitude, boundary radius, and nondecision time from (i) the mean resultant length and mean direction of response vectors and (ii) the mean and variance of response time (RT). We also present robust variants based on trimmed directional means and median/IQR timing summaries. Monte Carlo validation across a wide parameter range demonstrates accurate recovery and competitive prediction of joint RT–direction data, with orders-of-magnitude speedups relative to numerical likelihood methods. EZ-SDM makes spherical diffusion models practical for rapid screening, model comparison, and hierarchical Bayesian analysis.

Topic:

Mathematical models of psychological processes

Talks / 36

Contextual Probability Modelling of Decision under Ambiguity: Conditioning Schemes and Information-Theoretic Properties

Author: Mario Angelelli[1]

[1] *University of Salento*

Reasoning and decision-making processes may be affected by sources of uncertainty of different natures. Beyond stochastic uncertainty and known-risk scenarios, ambiguity may arise when the probability distribution over outcomes itself is partially inaccessible, as formalised by Ellsberg's urn models and the associated violations of the Sure-Thing Principle.

The present work proposes a probabilistic formulation of generalised Ellsberg-type decision models to investigate the emergence of ambiguity-related behaviours from partially accessible knowledge about contextual configurations underlying decisions. The model explicitly formalises uncertainty about such configurations as a discrete random variable, which is jointly distributed with the decision framing (winning condition specification) and the alternative available to the decision-maker. This construction extends Ellsberg's framework to mixtures of latent probabilistic scenarios. Joint probability distributions are derived from parameters describing the urn composition, providing a formal ground to distinguish between risky and ambiguous winning chances.

Formal results show how ambiguity-seeking and ambiguity-averse behaviours emerge from the same probabilistic structure under different conditioning schemes expressing distinct logic for decoupling contextual effects. The resulting probability distributions are also studied through information-theoretic quantities to measure the interaction among decision alternatives, winning conditions, and contextual configurations characterising the decision scenario. Specifically, the roles of risk proportion and the number of decision alternatives are examined via conditional mutual information and interaction information (or co-information), deriving conditions to assess the dominance of synergistic over redundant informational contributions in the admissible parameter domain. Counterexamples of model configurations violating these conditions are also presented through their interaction information.

More generally, the framework aims at providing a conceptual basis to relate ambiguity effects and bounded information resources. A complementary order-theoretic approach accounting for decision behaviours under ambiguity is also outlined and briefly discussed to encode accessibility relations between partial-knowledge descriptions and conditions for, or obstructions to, their combination and consistent extension.

Topic:

Mathematical models of psychological processes

Talks / 30

A query procedure for constructing maximally informative yet minimally long tests for skill assessment

Authors: Pasquale Anselmi[1]; Jürgen Heller[2]; Egidio Robusto[1]; Luca Stefanutti[1]

[1] *University of Padova*

[2] *University of Tübingen*

An assessment conducted within competence-based knowledge structure theory allows for identifying the latent set of skills an individual possesses, referred to as the "competence state", from the observed item responses. A recent approach to test development proposed within this framework

exploits concepts originally introduced in rough set theory to construct tests that are maximally informative about individuals' competence states (adding any item does not make the tests more informative) and minimal (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 possessing all of them is necessary to solve or endorse the item. Existing procedures for test construction assume that the particular sets of skills for which items are available or can be developed are known in advance. This assumption is easily satisfied when shortening an existing test because the relevant competencies are those already associated with the test items. By contrast, it may be unrealistic when constructing a test from scratch or improving an existing test because items requiring specific skills need to be developed, and some may be difficult or impossible to create. The talk presents a query procedure for test construction that is based on asking an expert, either a human expert or an AI expert, whether or not they can provide items requiring specific sets of skills. 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 once further queries would no longer lead to a more informative test. The resulting tests are maximally informative given the expert's capacity to provide items, while keeping the number of items either minimal or nearly minimal. The functioning of the procedure is illustrated using real-life examples.

Topic:

Knowledge structures

Talks / **3**

Partial comparability of latent traits across studies

Author: Wesam Asaad

Comparing latent traits such as self-esteem, cognitive ability, or stress across studies is difficult because differences in questionnaires, response scales, indicator content, and sampled populations can make numerically similar scores substantively non-equivalent. Apparent differences across studies may therefore reflect measurement heterogeneity rather than variation in a common underlying construct.

We develop a mathematical framework for partial comparability of latent traits across heterogeneous studies. Rather than imposing a single pooled latent representation, each study is allowed its own measurement system, while population-level constructs are characterized by the trade-off between study-specific fit and cross-study comparability. This yields a set of admissible population representations and recasts construct equivalence as a set-valued identification problem.

Our main theoretical result proves that, generically, the dimension of the admissible set equals the dimension of allowable measurement heterogeneity. Thus, under exact invariance, the set collapses to a single representation modulo standard rotational indeterminacy, whereas violations of invariance generate a higher-dimensional region of observationally equivalent constructs. A constructive algebraic example demonstrates how multiple incompatible population constructs can arise even in the population limit.

The framework also clarifies approximate invariance and alignment methods as procedures that select particular elements of a larger admissible set rather than recovering a uniquely defined construct. More broadly, it provides a principled basis for determining when cross-study latent trait comparisons are meaningful and when substantive conclusions are necessarily ambiguous.

Topic:

Psychometrics

Poster / 31

Kernel Density Estimation and the Overlapping Index : A Sensitivity Analysis Across Bandwidth Selectors and Kernel Functions

Authors: Giulia Calignano[1]; Ambra Perugini[1]; Massimiliano Pastore[1]

[1] *Università degli Studi di Padova*

How much does the choice of kernel function and bandwidth selector affect the nonparametric Overlapping Index $\eta = \int_{\mathbb{R}} \min\{f_A(x), f_B(x)\}\, dx$ (Pastore, and Calcagnì, 2019), when estimating distributional similarity between groups? Although η is theoretically distribution-free, its computation relies on kernel density estimation (KDE), introducing choices that may influence its properties (Sheather, and Jones, 1991; Silverman, 1986). This study presents a Monte Carlo analysis evaluating the impact of four kernel functions (Gaussian, Epanechnikov, triangular, cosine) crossed with five bandwidth selectors (nrd0, nrd, ucv, bcv, SJ) on the bias, precision, and stability of .

Data were generated from skewnormal distributions $\mathcal{SN}(\xi, \omega, \alpha)$, varying mean differences ($\delta \in \{0, 1, 2\}$), standard deviations ($\sigma \in \{1, 2\}$), skewness ($\alpha \in \{0, 5\}$), and sample sizes ($n \in \{30, 100, 200\}$), mirroring the simulation study comparing Cohen's d, CLES, η_p, and η by Perugini, Calignano, and Pastore (2026). Performance was evaluated through Relative Mean Bias (RMB) and Normalized Root Mean Square Error (NRMSE). Results indicate that the choice of KDE specification has a modest but non-negligible impact on $\hat{\eta}$: bandwidth selectors based on cross-validation (bcv) consistently yield the highest bias, whereas rule-of-thumb selectors (nrd0) combined with any kernel function produce the lowest RMB (0.039–0.041). Critically, all KDE configurations keep RMB within the acceptable ± 0.10 range across most scenarios, and RMSE decreases with sample size regardless of kernel choice. However, under strong skewness ($= 5$), large mean differences ($= 2$), and equal variances ($= 1$), bcv produces RMB values exceeding the ± 0.10 threshold (up to $+ 0.20$ for the triangular kernel), indicating that cross-validation bandwidth selection becomes unreliable under these conditions.

These findings suggest that η is robust to reasonable KDE choices and bandwidth selection matters more than kernel shape, and that nrd0 with any kernel, particularly triangular or Gaussian, represents a defensible default for applied use.

Topic:

Statistical methods

Talks / 1

Autocorrelated Sampling in Cognition: Implementing MCMC Models and Fitting Them Without Likelihoods

Authors: Lucas Castillo[1]; C. Stella Qian[1]; Adam N. Sanborn[1]

[1] *University of Warwick*

A key question in mathematical psychology is how people achieve sophisticated performance given limited resources: human behaviour is similar to the Bayesian ideal in many domains (Xu & Tenenbaum, 2007; Yuille & Kersten, 2006), yet given existing constraints optimality is impossible. Computer scientists have developed a way in which Bayes'rule can be approximated, a family of algorithms called Markov Chain Monte Carlo (MCMC; Brooks et al., 2011), which operates by drawing samples from the posterior distribution in an autocorrelated fashion. More recently, mathematical psychologists have shown that these inference algorithms describe human behaviour remarkably well in both low- and high-level cognition, from perception (Gershman et al., 2012) to numerical

estimation and judgments of probability (Dasgupta et al., 2017). Despite their success, their uptake has been sluggish—we believe because of the existing difficulty in their implementation. The algorithms are somewhat tricky to code, and should be implemented in a low-level programming language for speed. What is more, models using MCMC will lack an explicit likelihood function, and therefore model fitting will need to be carried out via simulation (Cranmer et al., 2020; Turner & Van Zandt, 2012). In this talk, we show how MCMC has been used to model different cognitive processes, and introduce attendees to the samplr R package (Castillo et al., 2025), which implements different MCMC algorithms as well as a cognitive process model incorporating them into a larger architecture. We also give an introduction as to how these MCMC-based models can be modelled using Approximate Bayesian Computation (Turner & Van Zandt, 2012), a simulation-based inference technique, focusing on which analyses are most diagnostic.

Topic:

Computational methods

Talks / 13

On race models for stop signal paradigms

Author: Hans Colonius[1]

Coauthors: Adele Diederich [1]; Paria Jahansa [1]

[1] *Oldenburg University*

The stop-signal task is a popular paradigm for investigating response inhibition: participants respond by pressing a button upon encountering a *go signal*; occasionally, a *stop signal* is presented after a variable stop-signal delay prompting participants to withhold their response. When the delay is long, participants typically fail to inhibit their response, resulting in a recorded reaction time. Conversely, for short delays, they often succeed in following the instruction, leading to no recorded response. *Stimulus-selective stopping* extends the standard stop signal task by occasionally presenting an 'ignore'signal instead of a stop signal, in which case participants are instructed to continue responding to the go signal.

Here we discuss some extensions of the classic race model (Logan & Cowan, *Psych. Review*, 1984) that take into account stimulus-selective stopping data and trigger failures. Using the concept of *(vine) copulas* to model stochastic dependency, we extend our previous results (Colonius, Jahansa, Diederich, *Comput. Brain & Behavior*, 2024).

Topic:

Mathematical models of psychological processes

Poster / 14

Modeling Expectations and Framing in Causal Judgment based on Contingency: A Bayesian Multinomial Processing Tree Approach

Author: Stefano Dalla Bona[1]

Coauthor: Michele Vicovaro [1]

[1] *Università degli Studi di Padova*

Causal judgment research has typically treated narrative cover stories in contingency learning tasks as neutral vehicles for conveying statistical information upon which people make decisions. We challenge this assumption, supporting that such narratives systematically shape causal inference through two mechanisms: (i) prior expectations about causal strength and (ii) deterministic versus probabilistic framing of the scenario.

Across two preregistered between-subjects experiments ($N = 195$ each), cover stories were designed to vary expectation strength and framing type, and were tested under both illusory null ($P = 0$; illusion of causality) and positive ($P = .50$) contingency conditions. Stronger positive expectations consistently increased causal endorsement, while deterministic framing reduced it by heightening sensitivity to disconfirmatory evidence. Crucially, these two factors operated independently, suggesting they reflect distinct cognitive processes.

To formally capture these effects, we developed and implemented a Bayesian Multinomial Processing Tree (MPT) model decomposing causal judgments into latent reasoning components: the probability of holding a positive expectation (E), adopting a deterministic interpretation (D), and endorsing causality under either an accepting (TA) or skeptical (TS) response mode.

The MPT model yielded three key findings: (1) expectations and framing exert independent, albeit additive, effects on causal judgment; (2) both response modes remain sensitive to contingency information, indicating that contextual factors modulate but do not override statistical learning; and (3) the illusion of causality under null contingency is amplified—but not fully determined—by an accepting response mode.

These results showed that cover stories are active elements of experimental design that shape causal inference in theoretically meaningful ways. More broadly, they illustrate the value of formal MPT modeling for disentangling the cognitive mechanisms underlying causal learning and the illusion of causality, with direct implications for how contingency learning paradigms are designed and interpreted.

Topic:

Models of cognition and learning

Talks / **33**

Large Language Models for the Theory-Driven Construction of Knowledge and Competence Structures

Authors: Debora de Chiusole[1]; Umberto Barbieri[2]

Coauthors: Luca Stefanutti [1]; Pasquale Anselmi [1]

[1] *University of Padua*

[2] *Pegaso Online University*

The methodological role of Large Language Models (LLMs) in supporting the theory-driven construction of knowledge and competence structures is examined within the framework of Competence-based Knowledge Structure Theory (CbKST). The work addresses a central challenge in CbKST, namely the development of cognitively meaningful and empirically valid structures that can support diagnostic and adaptive educational systems. To investigate this issue, we implemented a multi-agent system in which heterogeneous LLMs operate under different levels of autonomy in the generation and refinement of skill maps and competence structures. The resulting structures were empirically evaluated against the correctness of responses from elementary school students to arithmetic items. The findings suggest that LLM-generated structures are comparable to those developed by human experts when the generation process is guided by theoretical and structural constraints. In contrast, fully autonomous approaches tend to produce excessively granular and less parsimonious structures, limiting their interpretability and diagnostic usefulness. The study contributes to current discussions on the use of generative AI for interpretable, scalable, and empirically grounded modeling in educational assessment and adaptive learning systems.

Topic:

Knowledge structures

Keynote / 54

Classification of brain states that predicts future performance in visual tasks based on co-integration analysis of EEG data

Authors: Marie Levakova; Jeppe Høy Christensen; Susanne Ditlevsen[1]

[1] *University of Copenhagen*

Electroencephalogram (EEG) is a popular tool for studying brain activity. Numerous statistical techniques exist to enhance understanding of the complex dynamics underlying the EEG recordings. Inferring the functional network connectivity between EEG channels is of interest, and non-parametric inference methods are typically applied. We propose a fully parametric model-based approach via cointegration analysis. It not only estimates the network but also provides further insight through cointegration vectors, which characterize equilibrium states, and the corresponding loadings, which describe the mechanism of how the EEG dynamics is drawn to the equilibrium. We outline the estimation procedure in the context of EEG data, which faces specific challenges compared with the common econometric problems, for which cointegration analysis was originally conceived. In particular, the dimension is higher, typically around 64; there is usually access to repeated trials; and the data are artificially linearly dependent through the normalization done in EEG recordings. Finally, we illustrate the method on EEG data from a visual task experiment and show how brain states identified via cointegration analysis can be utilized in further investigations of determinants playing roles in sensory identifications.

Topic:

Computational methods

Talks / 27

Every response matters: Nested logit item response theory models for the development of smart item selection algorithms

Authors: Ottavia M. Epifania[1]; Andrea Brancaccio; Pasquale Anselmi; Egidio Robusto

[1] *Università di Trento*

In standard multiple-choice maximum-performance tests, responses are commonly coded dichotomously as correct or incorrect, collapsing all incorrect response options into a single category. This approach assumes that incorrect responses provide no information about an individual's latent trait level. This assumption is rather strong, as it implies that individuals either know the correct answer and respond correctly, or respond essentially at random when they do not know it. A more realistic assumption is that examinees often make educated guesses even when they are uncertain about the correct response. Several Item Response Theory (IRT) models have been proposed to account for this process, including the ability-based guessing framework (San Martín & De Boeck, 2006) and the IRT nested logit model (IRT-NLM; Suh & Bolt, 2010). Within the IRT-NLM framework, the selection among response options is conceptualized as a hierarchical sequential process. At the higher level, the process leading to the identification of the correct response is modeled using a standard dichotomous IRT model. At the lower level, conditional on an incorrect response, the selection of a specific distractor is modeled through the nominal response model (NRM; Bock, 1972). Unlike the standard

NRM, the dichotomous component of the IRT-NLM yields probability curves and information functions that can be directly compared with those obtained from conventional dichotomous IRT models. This allows for evaluating the additional information contributed by incorrect response options at an item-by-item basis. This contribution presents the results of simulation studies designed to evaluate the efficiency of adaptive item administration when item selection is based either on information functions derived from traditional dichotomous IRT models or on those obtained from the IRT-NLM. Specifically, we examine whether incorporating distractor-level information improves trait estimation accuracy and item-selection efficiency under different testing conditions.

Topic:

Psychometrics

Poster / 16

Stuck in the Web: How Psychological Network Analysis Traps Researchers in Its Own Structure

Authors: Hojjatollah Farahani[1]; Natasa Kovac[2]; Peter Watson[3]

[1] *Tarbiat Modares university*

[2] *Faculty of Applied Sciences, University of Donja Gorica*

[3] *MRC Cognition & Brain Sciences Unit , University of Cambridge*

Psychological network analysis has become a prominent framework for characterizing the interdependence among psychological variables, yet its methodological foundations are frequently undermined by conceptual and statistical weaknesses. This article examines how common practices in Gaussian graphical modeling and related network approaches introduce systematic distortions that compromise both inference and interpretation. A primary source of difficulty is the absence of well-defined system boundaries. Many published networks combine traits, states, symptoms, and composite scores within a single model, despite these constructs operating at different temporal scales and levels of abstraction. Such heterogeneity violates assumptions of granularity and comparability, and it obscures the meaning of estimated edges. Measurement issues further exacerbate these problems. Variability in reliability, conceptual redundancy among nodes, topological overlap, and shared method variance influences the covariance structure in ways that are rarely acknowledged, inflating edges and centrality indices while producing unstable community partitions.

Statistical estimation introduces additional vulnerabilities. Networks are often fitted to ordinal or highly skewed data using Pearson correlations, with limited attention to distributional assumptions or the appropriateness of regularization parameters in EBICglasso. Sample sizes are frequently inadequate relative to model dimensionality, and missing-data procedures are seldom justified, increasing the risk of unstable or non-identifiable solutions. Diagnostic tools which such as bootstrap confidence intervals, case-dropping stability analyses, and centrality stability coefficients which are inconsistently applied, leading to overconfidence in edge differences and node rankings. Cross-sectional networks are also routinely interpreted as if they reflected within-person dynamics or causal mechanisms, despite representing between-person differences.

These issues collectively reveal a pattern in which the method's visual appeal and intuitive metaphors encourage interpretations that exceed what the statistical model can justify. The article concludes with methodological recommendations aimed at strengthening the theoretical coherence, psychometric adequacy, and statistical robustness required for network analysis to function as a credible tool within mathematical psychology.

Topic:

Computational methods

Poster / 10

Algorithmic Counterfactual Explanations of Models as a Multi-Layer Framework for Artificial Psychology

Authors: Hojjatollah Farahani[1]; Natasa Kovac[2]

Coauthors: Mina Servati Samarin [1]; Nina Hubig [3]; Peter Watson [4]

[1] *Department of Psychology, Tarbiat Modares University, Tehran, Iran*

[2] *Faculty of Applied Sciences, University of Donja Gorica*

[3] *Interdisciplinary Transformation University Linz IT:U*

[4] *MRC Cognition & Brain Sciences Unit , University of Cambridge*

Two main methodological approaches have formed the basis of psychological science: a) correlational modeling and, lately, b) causal inference. Such details like observed associations and effect averages are not enough to answer the question: "What 'little fix'can influence changes in psychological outcomes in individuals?"

This paper proposes a psychologically inspired three-layer computational architecture for psychological simulations within the new and developing field of Artificial Psychology. The levels include nonlinear behavioral prediction (Random Forest), heterogeneous causal estimates (Causal Forest), and algorithmic counterfactuals identifying minimally sufficient paths of psychological change. R implementation serves as the basis for all analyses, preserving their transparency and reproducibility.

Random Forest delivers a stellar predictive performance (R2) while remaining non-causal. The Causal Forest analytical layer indicates significant heterogeneity of intervention effects, ranging from x to y, which demonstrates a high degree of individual variation in the psychological interventions'effectiveness. Using counterfactuals as input, the output logistics demonstrated sensible minimal-change trajectories based on the given causal structure.

Thus, "counterfactual outputs are not presented as alternative reality but as model-based simulations that are conditional to assumptions that are explicitly defined."The counterfactual outputs are not supposed to recreate past experiences that existing psychological mechanisms have processed but to assess the sensitivity of psychological outcomes to changes in the input feature vector.

Conceptually, the framework provides disambiguation on prediction, causal inference, and counterfactual reasoning as a basis for shifting predictive-only models to explanatory and intervention-centered models under Artificial Psychology. More broadly, machine learning plus counterfactual analysis "establish[es] a mathematically coherent computational viewpoint for mathematical psychology whose goal is not merely to infer what happened, but to develop a more basic question: Which minimum changes in a psychological structure of an individual can reroute the path for a certain set of outcome curves?

Topic:

Computational methods

Poster / 22

A Bayesian workflow for studying individual differences

Author: Anne Giacobello

Coauthor: Julia Haaf

Research on individual differences aims to recover stable, person-specific parameters from noisy experimental data. Hierarchical models (also called multilevel models) improve reliability through partial pooling, but they are typically evaluated mainly at the population level. Consequently, the quality of individual-level estimates often remains unchecked, leading to unstable individual estimates that produce attenuated correlations and conclusions that are driven more by modelling choices rather than true psychological variation. Therefore, evaluating models with inferences about individual differences must assess whether the model supports inferences about individuals. We propose a structured workflow for developing and evaluating Bayesian hierarchical models that treats validity of individual-level inferences as a main objective and is organised into three stages: (A) before fitting the model, (B) fitting the model, and (C) evaluating the (fitted) model.

A. Before fitting the model: We address defining the experimental design, selecting an initial model, and specifying informative priors for both population- and individual-level parameters. We also discuss prior-predictive checks to verify that the model can generate plausible heterogeneity and correlation structures.

B. Fitting the model: We focus on parameter recovery, computational diagnostics, and model validation. We discuss simulation-based recovery analyses, different computational diagnostics and model validation (such as cross validation), with a focus on recovery and validation of individual-level estimates.

C. Evaluating the model: From the wide range of options to evaluate the quality of a Bayesian model, we focus on posterior prediction, Bayesian fit indices, Bayesian model comparison using the Bayes factor, and sensitivity analyses. While many of the tools are well established, their integration into a workflow focused on recoverability, robustness, and interpretability at the individual level has, to the best of our knowledge, not yet been done.

Topic:

Computational methods

Talks / 37

Reference-based imputation for clustered longitudinal data

Author: Matthias Gondan

Coauthors: Lukas Legner ; Irene Alfarone

In clinical trials, patients may discontinue treatment or deviate from study protocol for various reasons. The assumptions made about these reasons inform the missing imputation strategy and directly impact the treatment estimate. In randomized controlled trials, reference-based imputation methods assume after a study dropout, patients may (partially) lose the advantage of the treatment arm, and their missing values are imputed using information from the reference arm.

In this work, we present two approaches from recent methodological contributions to reference-based imputation. The so-called information-anchored sensitivity analysis uses a Markov Chain Monte Carlo algorithm to generate the posterior distribution of the treatment effect. A frequentist method uses likelihood-based mixed models for repeated measures and estimates the variance by using bootstrap or jackknife.

We extend both methods to clustered longitudinal data, e.g. from group therapy studies. We investigate the confidence interval coverage and statistical power, as well as the impact of misspecifications of covariance structure on the treatment estimate.

Topic:

Statistical methods

Poster / 11

Specifying Meaningful Joint Hypotheses Across Studies: Bayesian Evidence Synthesis Revisited

Author: Laura Groot[1]

Coauthor: Daniel W. Heck [1]

[1] *Universität Marburg*

Bayesian evidence synthesis (BES) is a method for evaluating the empirical support for a common theory across a set of heterogeneous studies. Evidence is aggregated by multiplying Bayes factors from the individual studies, yielding the product Bayes factor (PBF). The PBF numerator aggregates evidence for the joint predicted hypothesis, which states that the study-specific predicted hypothesis holds in each study. This hypothesis reflects the belief that the theory holds across different contexts, treating each study-specific predicted hypothesis as an instance of the theory. However, since Bayes factors are relative measures of evidence, interpreting support for the joint predicted hypothesis via the PBF requires a clear understanding of the alternative against which it is compared. This joint alternative hypothesis arises "automatically" in the PBF denominator when multiplying the individual Bayes factors and states that the study-specific alternative hypothesis holds in each study. However, due to its implicit, bottom-up construction, this joint alternative is at risk of lacking a clear substantive or theoretical interpretation, which makes it unclear how results from BES can be understood. In this project, we investigate how different constellations of study-specific alternative hypotheses give rise to distinct joint alternative hypotheses. We focus on three common types of alternative hypotheses at the study-level: the null, the complement, and the unconstrained alternative. We develop principled recommendations for constructing meaningful joint alternative hypotheses that support clear interpretation of BES and are useful in the context of theory-testing.

Topic:

Statistical methods

Talks / 5

A Mixture Processing Account of Heavy-tailed Recall Error

Author: Amir Hosein Hadian Rasanan[1]

Coauthor: Nathan Evans [2]

[1] *University of Basel*

[2] *University of Liverpool*

Visual working memory (VWM) is a system responsible for storing limited visual information for a short period of time and serves as the interface between the visual system and high-level cognitive processes such as visual attention and decision making. Therefore, understanding the cognitive constructs of VWM is fundamental to understanding other higher-level cognitive processes. Although several measurement models have been proposed to disentangle guess responses from memory-based responses, the process underlying guess responses remains unclear. Moreover, existing measurement theories primarily focus on the distribution of response errors and overlook response-time behavior. In this project, we proposed a mixture circular diffusion model that can simultaneously account for response-time and recall error distributions and predict the heavy-tailed distribution of recall errors, a key signature of guess responses. The core assumption of the model is that the noise in evidence accumulation during the guessing and memory-based recall processes can differ. To test this assumption, we first show, through an extensive simulation study, that the relative noise of the guessing and recall processes can be reliably estimated from response time and response error data. Second, by reanalyzing six datasets with a total of $N = 421$ subjects, we show that the model successfully captures qualitative patterns in mean response time and mean absolute error,

and achieves high precision in predicting response-time and response-error distributions. Moreover, based on both quantitative and qualitative model comparisons, the proposed model explains behavior better than the other competing models considered. Overall, this work provides a deeper insight into the cognitive constructs of VWM and the sampling process from memory.

Topic:

Models of cognition and learning

Talks / 20

Building precedence relations from data: A theoretical perspective on an empirical approach

Author: Jürgen Heller[1]

[1] *University of Tübingen*

Precedence relations characterize dependencies between the mastery of different items within a knowledge domain in a direct and intuitive way. Moreover, the one-to-one correspondence between precedence relations and quasi ordinal knowledge spaces forms a cornerstone of knowledge structure theory and enables the construction of knowledge structures from empirical data. Various ad hoc approaches for this task have been proposed, which are based on analyzing the 2x2 contingency tables induced by pairs of items. The talk views precedence relations from a theoretical perspective within the framework of probabilistic knowledge structures. Contingency tables are interpreted as realizations of probabilistic projections, and their properties (such as stochastic independence) are analyzed to derive a characterization of precedence relations that can provide a theoretical foundation for their empirical identification.

Topic:

Knowledge structures

Talks / 12

Critically testing the drift diffusion model as an account for response times in risky decision making

Author: Sebastian Hellmann[1]

Coauthor: Thorsten Pachur [1]

[1] *TU Munich*

The popular drift diffusion model (DDM) predicts an inverse U-shaped relationship between choice strength and response times (RTs). Here, we provide a large-scale test of this relationship in risky choice, a domain in which it has not been systematically evaluated. We compiled 14 publicly available datasets (1,388 participants, 199,157 choices in total), spanning a wide range of experimental paradigms. We fitted a hierarchical model that integrates cumulative prospect theory (CPT) with DDM-style evidence accumulation (CPT-DDM), allowing for joint modeling of choices and RTs. We find substantial heterogeneity across datasets: while the CPT-DDM captures overall RT distributions well, the predicted U-shaped relationship between choice strength and RT is weak or absent in several tasks, particularly in simple accept–reject paradigms. In contrast, more complex lottery choices show patterns broadly consistent with DDM predictions. Importantly, parameter estimates

from the CPT-DDM closely align with those obtained from a standard CPT model fitted to choices only, indicating that inferences about underlying preferences remain robust even when RTs are not incorporated. Together, these findings suggest that while the DDM provides a useful approximation of decision dynamics in some risky choice settings, its applicability depends critically on task characteristics and may require extensions that account for additional processes such as information search.

Topic:

Mathematical models of psychological processes

Talks / 53

Optimal designs for Thurstonian IRT models based on metric paired comparisons

Author: Heinz Holling[1]

[1] *University of Münster*

Thurstonian IRT models represent a milestone in trait measurement, as they enable the estimation of interindividual trait values based on ipsative measurements, particularly pairwise comparisons. An important issue for the practical application of these models is the development of optimal designs for them. Optimal designs of item pairs are characterized by combinations of factor loadings that optimize specified criteria, such as the correlation between the estimated and true trait values or the volume of the confidence ellipsoid of the trait values. For many applications, such as personnel selection, paired comparisons should consist of equally weighted items. This condition requires the development of novel optimal designs. In addition to the properties of the optimal designs developed in the literature to date, two further requirements must be given special consideration: (a) the restriction of the design space and (b) the condition that the alternatives must load on mutually distinct factors. In this presentation, solutions to the problem of optimal design will be presented that significantly outperform the methods currently known in the literature in terms of accuracy and the number of required pairwise comparisons.

Topic:

Measurement and scaling

Talks / 15

An algebraic solution to Marley's problem on four alternatives

Author: Karim Kilani[1]

Coauthor: Hans Colonius [2]

[1] *LIRSA, Conservatoire National des Arts et Métiers, Paris*

[2] *University of Oldenburg*

We study Marley's best–worst choice problem in the case of four alternatives. The problem consists in characterizing systems of best–worst choice probabilities that admit a representation through a probability distribution over rankings. Although previous work established geometric descriptions of the associated polytope, a fully explicit constructive characterization was still missing.

We develop an algebraic solution based on a reformulation of the compatibility equations in terms of adjacent swaps between rankings. Using the Johnson–Trotter ordering of permutations, we show that the system possesses a recursive structure allowing all ranking probabilities to be expressed explicitly from a single free probability. The construction propagates recursively along the Johnson–Trotter chain and reveals a simple alternating structure linking adjacent rankings.

This approach yields a complete characterization of admissible best–worst probabilities. In particular, we derive a system of 144 linear inequalities that are necessary and sufficient for the existence of a compatible ranking distribution. These inequalities arise naturally as sums of two ranking probabilities associated with permutations of opposite parity in the Johnson–Trotter ordering, and recover exactly the number of facet-defining inequalities previously obtained through polyhedral methods.

Whenever these conditions are satisfied, all ranking probabilities are fully determined from a single free probability. The results show how combinatorial structures on permutations can lead to explicit solutions for probability distributions over rankings, and suggest that this approach may provide a promising route toward solving the general Marley problem, which remains open for an arbitrary number of alternatives.

Topic:

Mathematical models of psychological processes

Talks / 8

Accounting for Censored Data in Horse-Race Models of the Stop-Signal Paradigm

Author: Lars Kulbe[1]

Coauthors: Andrew Heathcote [2]; Zachary L. Howard [3]; Dora Matzke [1]

[1] *University of Amsterdam*

[2] *University of Newcastle; University of Amsterdam*

[3] *The University of Western Australia; Defence Science and Technology Group, Canberra, Australia*

The stop-signal task (SST) is a popular paradigm for investigating response inhibition. Because successfully inhibited trials run until the response deadline expires, shorter deadlines reduce experiment duration and help maintain participant engagement. Short response deadlines additionally enforce fast responses and discourage strategic slowing beyond what experimental instructions alone achieve, and may improve SST performance in some clinical populations. Such deadlines, however, risk losing slow but valid responses to upper censoring, which is known to distort parameter estimates derived from response-time distributions unless properly accounted for. We formally introduce and validate an extension of previous Bayesian parametric models of the SST that accounts for upper response censoring via an adjustment to the likelihood function. We show via simulation-based calibration that our censored estimation procedure is well calibrated. Individual-subject parameter recovery at near-asymptotic trial counts yields unbiased estimates across a wide range of censoring levels, whereas existing models that do not account for censoring show progressively biased estimates as the level of censoring increases. Residual small-sample bias in individual-subject parameters at realistic trial counts reflects a coherent prior-likelihood tradeoff for stop parameters that are only weakly informed by the data. Hierarchical estimation with realistic trial counts compensates for this lack of information via partial pooling and shows unbiased recovery of population-level means, with largely unbiased population variances. Our approach highlights the importance of accounting for slow but valid responses in the SST and provides a model that accommodates them. We provide a readily available implementation of this model in the EMC2 R package, enabling researchers to employ strict response deadlines strategically, or as a principled post-hoc procedure for the exclusion of slow outliers. We emphasize the inferential risks of misspecified models that do not account for upper response censoring.

Topic:

Bayesian models

Talks / 29

Stochastic dominance in canonical response time models

Authors: Lukas Legner[1]; Matthias Gondan[1]

[1] *University of Innsbruck*

Many studies in cognitive psychology measure task performance through response time and/or accuracy. The well-known speed-accuracy tradeoff directly links these two measurements through participants'strategy and prioritization. However, the usual practice is to either directly remove participants with non-ceiling accuracy or exclude wrong responses from their analysis model altogether. The seminal work of Townsend and Nozawa (1995, *Journal of Mathematical Psychology*) on two-factorial experimental designs studied interaction contrasts of response time distributions and made predictions for cognitive architectures including serial and parallel processing, with exhaustive and self-terminating stopping rules. This non-parametric method involves two assumptions, selective influence (i.e., manipulation of one factor may not influence the processing time of the other) and stochastic dominance (i.e., weaker stimulus salience leads to slower processing).

We have recently published a generalization of these model predictions to both correct and incorrect responses (Gondan & Legner, 2026, *Journal of Mathematical Psychology*). To do so, we extended the stochastic dominance assumption to ordered survivor functions, ordered subdistributions of correct processes, and ordered subdensities of incorrect processes. We show that these assumptions are met by canonical models of response time in two-choice models; the Poisson race model, the diffusion model, the linear ballistic accumulators and the Poisson random walk model under "reasonable" parameter restrictions. We use the term reasonable to say that evidence accumulation is overall positive, the starting point is not biased towards the wrong response, and greater salience promotes correct decisions and reduces errors. These models naturally account for speed-accuracy trade-offs at the stage of evidence accumulation. Therefore, our extension can be readily applied to many experimental designs in cognitive psychology.

Topic:

Mathematical models of psychological processes

Talks / 19

A new geometric model for Best–Worst choice

Authors: Adele Diederich[1]; Keivan Mallahi Karai[2]; Majid Salamat

[1] *University of Oldenburg*

[2] *Constructor University*

We introduce a stochastic-geometric model for best–worst decision making with three alternatives. Best–Worst tasks require a decision maker to identify both the most preferred and the least preferred option, thereby producing a complete ranking rather than a single choice. The proposed model—the hexagonal model—is motivated by a number of other geometric approaches to multi-alternative decision making, in particular the Cube Model and the Disk Model.

We also develop an approximation schemes for estimating the parameters of the model and compare the results with some existing Best–Worst models.

This talk is based on a joint work with Adele Diederich and Majid Salamat.

Topic:

Mathematical models of psychological processes

Poster / 34

Validated Multiverse Meta-Analysis: Expert-Constrained Specification Spaces and Post-Selection Inference

Author: Matteo Manente[1]

Coauthors: Filippo Gambarota [1]; Livio Finos [1]; Gianmarco Altoè [1]

[1] *University of Padova*

Meta-analytic conclusions are often sensitive to researcher degrees of freedom, including study selection, outcome operationalization, effect-size computation, and model specification. Multiverse Analysis addresses this problem by systematically assessing alternative analytic pathways, but most implementations remain descriptive and provide limited inferential control.

This project proposes a formal framework for inferential Multiverse Meta-Analysis that integrates structured expert elicitation with post-selection inference. Expert elicitation is used to identify and validate reasonable analytic decision nodes, thereby constraining the specification space to theoretically and methodologically admissible pathways. The resulting multiverse is then analyzed using Post-Selection Inference for Multiverse Analysis (PIMA), a permutation-based procedure that controls the Family-Wise Error Rate across correlated specifications through max-t adjustment.

The framework conceptualizes multiverse meta-analysis as a constrained model-space inference problem, allowing valid statistical inference across multiple dependent analytic specifications. Particular attention is given to dependence structures among pathways, multiplicity correction, and robustness of conclusions under analytic uncertainty.

In future applications, the framework will be applied to meta-analyses of Eye Movement Desensitization and Reprocessing (EMDR) therapy for Post-Traumatic Stress Disorder (PTSD), a literature characterized by substantial analytic heterogeneity and conflicting conclusions regarding treatment efficacy.

Topic:

Statistical methods

Poster / 49

A Dynamical Systems Model of Emotion Regulation Ability Across Psychological Context Space

Author: Filip Melinscak[1]

Coauthors: Daniel Reiter [1]; Frank Scharnowski [1]

[1] *University of Vienna*

Adaptive emotion regulation is essential for coping with changing environmental demands, yet regulation ability is often conceptualized as a relatively stable trait. Existing theoretical work has independently emphasized the importance of temporal dynamics in emotional processing and similarity-based generalization across psychologically related contexts. However, these principles have not yet been integrated into formal models of emotion regulation ability.

Here we introduce a dynamical systems model in which emotion regulation ability is represented not as a single scalar quantity, but as a landscape defined over a latent psychological similarity space of contexts. The model couples three interacting processes that evolve in continuous time: emotional intensity driven by environmental input, regulation that reduces emotional activation but depends on local regulation ability, and adaptive changes in regulation ability arising from prior successful regulation. Learning remains localized around experienced contexts while simultaneously generalizing to nearby contexts according to a Shepard-like exponential similarity gradient. Consequently, regulation ability evolves as a dynamic surface across psychological contexts rather than as a globally uniform trait.

Preliminary simulations suggest that the model captures several qualitative features of adaptive regulation. Repeated regulation within a given context produces localized increases in regulation ability and reduces subsequent emotional responses. At the same time, these gains partially transfer to similar contexts, with generalization decreasing systematically as psychological distance increases. The model further generates differentiated "islands"of regulation ability across context space, suggesting a potential mechanism through which adaptive specialization and fragmented regulation patterns emerge over time.

Ongoing work investigates how trajectories of emotional experience shape evolving regulation landscapes, including the effects of context sequences, individual differences in generalization width and learning rate, and the conditions under which regulation ability becomes broadly generalized rather than narrowly context-bound.

Topic:

Mathematical models of psychological processes

Poster / 48

Meta-learning approaches for dynamic structural equation models

Author: Jan Luca Neduchal[1]

Coauthors: Holger Brandt [1]; Kento Okuyama [1]

[1] *University of Tübingen*

Dynamic structural equation models (DSEM) are a popular framework for analyzing intensive longitudinal data. They combine time-series modeling with structural equation modeling. However, model evaluation remains difficult: traditional model fit indices become unreliable in complex settings, and appropriate modification indices that identify measurement model misspecifications do not exist.

This work presents three ensemble methods—linear stacking, hierarchical stacking, and model orthogonalization—that capture and account for time-varying factor loading shifts in the measurement model that would otherwise remain hidden. The methods stack time-specific likelihoods from an ensemble of base learners, converged DSEM specifications, that solely differ in their time-invariant factor loadings. By increasing and decreasing base learner weights according to their local fits, the proposed stacking approaches produce interpretable ensemble weights and better calibrated DSEM

with respect to their latent variable estimates and time-specific likelihoods. The ensemble learners shed light on local misspecifications and efficiently adapt to shifts in the measurement model.

Results from a simulation study suggest that all three methods correctly identify the time point at which the shift occurrs and adjustes their loading matrix in the following period. Hierarchical stacking and model orthogonalization use prior information from linear stacking and produce more likely predictions and less biased estimates. Further, manually accounting for shifting factor loadings through a weight-informed DSEM provides the best results among all learners.

This offers a practical view on detecting and accommodating misspecifications solely in the measurement model, leaving the structural dynamics unchanged until dedicated DSEM fit diagnostics become available.

Topic:

Psychometrics

Talks / 47

Variational dynamic latent class analysis

Author: Jan Luca Neduchal[1]

Coauthor: Augustin Kelava [1]

[1] *University of Tübingen*

Dynamic latent class analysis (DLCA) captures within-person dynamics across qualitatively discrete latent states. In each state, these dynamics are modeled using a dynamic structural equation model. Transition probabilities between the discrete latent states are described by a Markov-switching model.

While this flexible framework is particularly attractive for intensive longitudinal data in psychological research, it poses substantial inferential challenges. Existing Bayesian implementations rely on Markov-chain Monte Carlo sampling, whose computational cost quickly becomes prohibitive as model complexity grows.

To address this limiting factor, we develop a structured variational approximation for single-level DLCA. Standard mean-field variational inference imposes a factorization between the discrete latent states and the continuous factor scores, thereby breaking the central dependence that defines DLCA: the coupling between the states and the factor scores they generate. Our structured approximation explicitly preserves this coupling, as well as the Markov properties of both latent sequences. For parameters shared across persons, we impose a factorized variational distribution estimated via coordinate ascent variational inference. Person-specific parameters are jointly approximated using a forward-backward algorithm in the variational E-step. This yields a two-step estimation procedure that mirrors the EM algorithm for regime-switching state-space models while preserving a Bayesian interpretation.

We conduct a simulation study to assess the performance of the proposed structured variational approximation in comparison with existing MCMC-based implementations of DLCA. Our results indicate that the variational approach performs comparably to MCMC while reducing the computational cost substantially, making the Bayesian estimation of DLCA feasible for larger samples and longer time series typical of intensive longitudinal data in psychological research.

Topic:

Psychometrics

Talks / 4

On the modeling of local dependence

Author: Stefano Noventa[1]

Coauthors: Andrea Spoto [1]; Augustin Kelava [2]; Jürgen Heller [3]

[1] *University of Padova, Department of General Psychology*

[2] *Universität Tübingen, Methods Center*

[3] *Universität Tübingen, Department of Psychology*

Violations of the assumption of local independence are a fundamental issue in Item Response Theory as they threaten model validity and bias the parameter estimates. For such a reason, a plethora of tests and approaches has been devised in the last forty years to detect or to model such violations. Nonetheless, local dependence remains an open problem, with somewhat blurred boundaries due to the lack of a general framework for dealing with the different notions of dependence that have been suggested in literature. The present contribution has a two-fold aim: On the one hand, to review and collect some of the approaches available in the literature; on the other hand, to suggest a possible systematization of some existing and some new approaches to local dependence by following a unified perspective on assessment models that merges Knowledge Structure Theory, Item Response Theory, and Cognitive Diagnostic Assessment. As a result, the two primitive notions of structure and process introduced by the unified framework allow to formalize and discuss deterministic and probabilistic modeling mechanisms of local dependence.

Topic:

Psychometrics

Poster / 17

STRUCTURA: A Bayesian Network Framework for Information-Theoretic Scoring and Behavioral Anomaly Detection

Authors: Matteo Orsoni[1]; Mariagrazia Benassi[1]; Giovanni Briganti[2]; Marco Scutari[3]

[1] *Department of Psychology, University of Bologna*

[2] *University of Mons*

[3] *Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA)*

Network psychometrics redefines psychological constructs as systems of interacting components, where item interdependencies constitute the construct itself. No measurement framework has yet assessed whether individual response patterns are structurally coherent (that is, consistent with the conditional dependencies encoded in the construct's network) while also detecting item-level anomalies and establishing normative standards that respect that network structure.

We introduce STRUCTURA, a Bayesian network (BN) framework for information-theoretic scoring and behavioural anomaly detection. From normative data, a directed acyclic graph is learned and compiled into a junction tree for exact inference. Three metrics are defined using the Markov blanket as local evidence: 1. the Topological Expected Success Score (TESS) that sums conditional success probabilities, yielding a structurally weighted ability estimate; 2. the Global Self-Information (SI) that quantifies the informational atypicality of the full response profile, mapped onto a normative centile scale; 3. the Local SI that provides a signed per-item anomaly metric, distinguishing proxies of careless errors and lucky guesses, with these anomalies flagged at the 5th and 95th normative percentiles, respectively.

We apply STRUCTURA to Raven's Coloured Progressive Matrices. In one example, a Guttman inversion pattern (failing easy items but succeeding on hard ones) gets 57.7% raw accuracy. However, the TESS was 52.9%, yielding a discrepancy attributable to structural incoherence in the response pattern. Global SI shows the profile is unusual, reaching the 98.5th centile. Local SI highlights 3 careless errors and 7 lucky guesses during the test.

STRUCTURA uses the structure of the construct to refine an information-theoretic scoring process. This creates a clear, and explainable quantitative framework that matches network psychometrics' logic and quantifies its uncertainties. For trait-like constructs, it can provide a sound normative approximation. For dynamic states, it serves as a starting point for progressive idiographic refinement through hierarchical Bayesian updating.

Topic:

Measurement and scaling

Poster / 28

Lessons from and for metatheory: Can we tell rule- and similarity-based grammar learning apart?

Author: Moulshree Rana[1]

Coauthors: Mark Blokpoel [1]; Olivia Guest [1]; Iris van Rooij [2]

[1] *Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; Department of Cognitive Science and Artificial Intelligence, Radboud University, Nijmegen, The Netherlands*

[2] *Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; Department of Cognitive Science and Artificial Intelligence, Radboud University, Nijmegen, The Netherlands; Department of Linguistics, Cognitive Science, and Semiotics, Aarhus University, Denmark*

In grammar learning, a distinction is drawn between a learning process involving the extraction of abstract rules (termed rule-based process) and a process involving computing similarity between current and past instances (termed similarity-based process). Several empirical criteria have been proposed to identify the underlying process in lab-based tasks of artificial and real grammar learning. However, empirical attempts have been unable to clearly demarcate the two processes. We aim to examine the reasons behind this incapability. We adopt a mixture of theoretical (van Rooij & Baggio, 2021) and metatheoretical approaches (Guest & Martin, 2023). First, we formalise verbal descriptions of the two purported processes as presented by psycholinguists in studies involving grammaticality judgements. By ensuring that the formal theories meet the criteria of transparency, sufficiency, unambiguity, and possibility (van de Braak et al., 2026) assumptions hidden in the verbal theory are unveiled and commitments are imposed. Second, formal theories of this calibre afford precise computer simulations. These simulations are used as a theoretical tool to explore the properties, assumptions and commitments of the formal theory (Blokpoel & Guest, in prep). From our observations, we draw out metatheoretical lessons at different levels: (specific) for accounts of grammar learning and grammaticality judgements, (domain) for domains in cognition featuring explanations in the form of dichotomies, (theoretic) for the development and assessment of theories of cognitive capacities generally, and (meta-theoretic) for the further development of metatheoretical methods. Properly formalising rule- and similarity-based theories in this manner demonstrates a path for psychological researchers to walk away from explanations pitting two theoretical accounts against each other presented as dichotomies. By focusing instead on the explanatory claims for each theory, we can overcome the unrealistic framing of "one camp winning out"that has driven endless production of several similar experiments in the hope of declaring a "winner".

Topic:

Mathematical models of psychological processes

Talks / **32**

Issues with the M-ratio as measure of metacognitive efficiency

Author: Manuel Rausch[1]

[1] *Alpe-Adria-Universität Klagenfurt*

The m-ratio has become the de facto standard for quantifying metacognitive ability, with a substantial portion of recent research on metacognition relying on this measure. The m-ratio has become the method of choice because it has been argued to measure metacognitive ability while controlling for discrimination performance, discrimination criteria, and confidence criteria, even without the explicit assumption of a specific generative model underlying confidence judgments. However, the theoretical and statistical foundations of the m-ratio have increasingly come under scrutiny. Previous studies have shown that the m-ratio neglects the dynamics of the decision process and that its reliability is low unless the number of trials is quite large. Here, I show that in contrast to a common assumption, the m-ratio is not free from assumptions about the generative model underlying confidence. Instead, the m-ratio implicitly assumes a generative model of confidence according to which the evidence underlying confidence judgments is sampled independently from the evidence used in the discrimination decision process, from a Gaussian distribution truncated at the discrimination criterion. Comparing the model underlying the m-ratio to ten alternative models of confidence and metacognition derived from signal detection theory, I reanalyzed 11 previously published experiments and two previously unpublished experiments. The results revealed that the m-ratio model was consistently outperformed by at least one alternative model. I also present the statConfR package, which allows researchers to test the assumptions underlying the m-ratio. Overall, I argue that the field's widespread reliance on the m-ratio is misplaced

Topic:

Mathematical models of psychological processes

Talks / **2**

You may not be so powerless after all

Author: Michel Regenwetter[1]

Coauthors: Marc Jekel [2]; Meichai Chen [1]

[1] *University of Illinois Urbana-Champaign*

[2] *Uni Koeln*

According to common wisdom, analyzing multiple data sets statistically goes hand in hand with either deteriorating power or explosive growth in sample size. This makes it difficult to study multiple different effects. We consider one-sided tests about proportions as a case in point. We turn conventional wisdom on its head by viewing a theory's predictions as a logical conjunction of hypotheses. Casting either a conjunction of Null hypotheses, or a conjunction of Alternative hypotheses, as a single prediction, namely an order-constrained model, leads to a rapid gain in parsimony, not an explosion of model complexity. Using order-constrained inference, we show that power increases rapidly with the number of component hypotheses. Likewise, in Bayesian order-constrained inference the strength of evidence for or against a model, relative to a competitor accumulates dramatically with the number of component hypotheses. As a result, combining many predictions buys us much Bayes factor power to accumulate strong evidence for or against models, even for strikingly small sample sizes.

Topic:

Other (please comment above)

Poster / **43**

Eye Riders Assessment (ER-A), a novel application to measure sustained attention using Hierarchical Drift Diffusion Model Approach

Authors: Alice Riccardi[1]; Matteo Orsoni[1]; Sara Magri[2]; Mariagrazia Benassi[1]

[1] *Department of Psychology, University of Bologna*

[2] *Develop-Players*

Gamified cognitive assessment has emerged as a potentially viable approach for ecologically valid assessment of sustained attention. However, conventional performance metrics such as mean accuracy and reaction time are inadequate in distinguishing the discrete cognitive processes underlying behavioural change.

The present study examines two versions of Eye-Riders A, a gamified sustained attention task modeled on an endless-runner paradigm, in which participants execute jump or slide responses to sequentially presented obstacles. A short and a long version were administered to 125 participants in a counterbalanced within-subjects design. Preliminary ANOVA confirmed that the two versions are matched in overall difficulty ($F(1, 123) = 0.051$, $p = .821$, $^2 < .001$) yet revealed a large Level × Version interaction ($F(1, 123) = 120.3$, $p < .001$, $^2 = .109$). This finding could be interpreted as indicative of a robust practice effect. No difference in overall performance was observed between versions ($F(1, 123) = 1.013$, $p = .316$), confirming successful counterbalancing.

These findings reveal a fundamental limitation of aggregate measures: they are unable to determine (1) whether the practice effect reflects a genuine improvement in attentional processing efficiency rather than a strategic relaxation of response thresholds; (2) whether a within-session vigilance decrement is present in both versions and differs in magnitude or trajectory between them; and (3) whether the two versions diverge at the level of latent cognitive parameters despite equivalent surface performance, thereby informing selection of the optimal version for sustained attention assessment.

To address these points, a Hierarchical Drift Diffusion Model is applied to decompose performance into drift rate, boundary separation, and non-decision time, estimated as a function of task version, temporal block within session, and administration order, with by-subject random effects. This framework applied to Eye Riders A would provide a process-level account of learning and vigilance dynamics.

Topic:

Psychometrics

Talks / **41**

The Noisy Comparison Theory: A Context-Sensitive Model for Decision Making Under Risk

Authors: Melvin Marti; Sebastian Olschewski; Jörg Rieskamp

People often make decision under risks, such as choosing between means of transportation, medical treatments or life insurances. Traditional economic theories of choice assume that people evaluate the available choice options independently of each other. However, these models cannot explain

how the context affects people's choices. Therefore we propose the Noisy Comparison (NC) theory according to which people compare the outcomes of options with each other. When performing these comparisons the theory assumes that people devote limited sensitivity to the probabilities of the outcomes and their subjective value (i.e. utilities). According to the theory risky choices should depend on the context, so that features of the options'outcome distributions should affect people's decisions. The suggested theory is rigorously tested against standard decision theories in a new experiment and previous studies. The theory is able to predict context dependencies across dependent and independent outcome structures: Changes in the outcome distributions and the presentation of outcomes (coalesced or split) systematically shifted choice proportions for skewed gambles, which alternative theories cannot capture. The re-analyses of two existing datasets demonstrates that the NC theory can account for a broad set of classical decision effects. The NC theory is able to predict these context effects by assuming a cognitive processes that interact with the presentation format of joint outcome distributions. A quantitative model comparison shows that the NC theory has a higher predictive accuracy in comparison to standard decision theories and approaches the predictive accuracy of highly-flexible context-dependent neural networks. In sum, the NC theory provides a parsimonious, theory driven, cognitively plausible, and descriptively accurate model of decision making under risk.

Topic:

Models of cognition and learning

Talks / 52

Knowledge Space Theory as a Method for Curriculum Learning in Machine Learning

Authors: Milan Segedinac[1]; Goran Savić[1]; Peter Steiner[2]; Andrej Hložan[1]; Matija Matović[1]; Marko Njegomir[1]; Mihaela Osmajić[1]

[1] *University of Novi Sad*

[2] *St. Gallen Universtiy of Teacher Education/ ETH Zurich*

Both human learners and artificial learners appear to benefit from structured learning trajectories, yet the relationship between these two forms of learning is still insufficiently understood. This paper investigates whether Knowledge Space Theory (KST), a framework originally developed to model human learning, can provide a theoretically grounded method for constructing curricula for curriculum learning (CL).

CL is a training paradigm for machine learning (ML) models in which the training data is organized along a predefined curriculum that has gained increasing attention recently (Wang et. al., 2021). KST represents learners'mastery within knowledge states structured by a surmise relation, which captures prerequisite dependencies among item types. Thereby, feasible learning paths, which can be interpreted as a curriculum, are identified (Doignon & Falmagne, 2012). This parallel motivates the following research question: Can a KST-based learning path yield a curriculum that improves training outcomes compared to standard training in CL?

To answer this question, 500 neural networks were trained on different numbers of geometric image classification tasks. Each network's per-task performance after training was interpreted as a response pattern of an artificial learner. Inductive Item Tree Analysis (IITA), a data-driven method for inferring surmise relations from response patterns, was applied to identify the surmise relation among the tasks. The resulting surmise relation was used to organize the training items into hierarchical curriculum tiers for CL.

The KST-based curriculum outperformed standard training without CL, increasing validation accuracy from 0.81 to 0.87, whereas reversing the derived curriculum led to the weakest performance (0.79). These findings suggest that KST can capture non-obvious dependencies in neural network learning and may serve as a viable basis for curriculum design in machine learning. At the same

time, the study remains limited to a small synthetic domain, highlighting the need for broader validation in more complex settings.

Topic:

Knowledge structures

Poster / 23

Are we getting interactions wrong? The role of link functions in psychological research

Author: Laura Sità[1]

Coauthors: Enrico Toffalini ; Filippo Gambarota ; Margherita Calderan ; Tommaso Feraco

[1] *University of Padova*

Psychology researchers frequently test interaction effects, often to identify moderators of known main effects, but also because interaction testing has become a routine extension of main-effect analyses. These analyses typically rely on classical linear models or ANOVA, both of which default to identity link functions. Yet many psychological outcome variables are bounded or strongly non-normally distributed (e.g., accuracies, response times, sum scores, error counts, proportions) and psychometric variables more broadly are well known to deviate from normality (Micceri, 1989). When the underlying distribution or link function is misspecified, interaction estimates can become biased or even reversed (Domingue et al., 2022), leading to results that appear meaningful but derive from model misspecification. Link functions, in particular, are crucial, because they transform equal intervals on one scale into different intervals in another, which is fundamental to what interactions represent. Despite these risks, our field lacks quantitative evidence on how frequently researchers (a) use appropriate link functions, (b) explicitly report them and (c) obtain significant interactions under potentially misspecified models. This project combines a pre-registered systematic review, illustrative simulation studies demonstrating how identity link misspecification can yield false-positive interactions for outcomes such as accuracy, response times and questionnaire sum scores. Finally, it aims to develop practical guidance to help researchers make appropriate link-function choices and avoid model misspecification in everyday practice.

References:
Domingue, B. W., Kanopka, K., Trejo, S., Rhemtulla, M., & Tucker-Drob, E. M. (2022). Ubiquitous bias and false discovery due to model misspecification in analysis of statistical interactions: The role of the outcome's distribution and metric properties. Psychological Methods, 29(6), 1164–1179.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105(1), 156.

Topic:

Statistical methods

Talks / 35

Characterizing and Breaking Gradedness in Polytomous Knowledge Structures

Authors: Andrea Spoto[1]; Luca Stefanutti[1]

[1] *University of Padova*

Gradedness plays a central role in the identifiability of knowledge structures. In dichotomous Knowledge Space Theory, its main properties and structural implications are well established, whereas the polytomous case appears substantially more complex. This work investigates gradedness in polytomous knowledge structures and provides structural conditions for both its characterization and its failure.

A distinction is introduced between weak and strong forms of gradedness, together with necessary and sufficient conditions under which gradedness fails in a given item. Stronger sufficient conditions breaking gradedness are then examined. The concept of bijectively related items is introduced as a generalization of equally informative items from classical Knowledge Space Theory. Several of the conditions eliminating gradedness arise in psychologically meaningful situations, including reverse-scored items, monotone transformations between response scales, and items sharing equivalent informational content.

Gradedness is known to induce non-identifiability issues in polytomous knowledge structures, whereas several of the structural conditions considered here are incompatible with gradedness. Bijectively related items are further studied through generalized discriminative reductions of polytomous structures.

The proposed results extend classical notions from dichotomous Knowledge Space Theory to the polytomous setting and clarify the structural role of gradedness in identifiability problems. Moreover, the results may contribute to the development of diagnostic criteria for detecting gradedness and identifiability in polytomous structures, and suggest principled ways of constructing non-graded structures that remain psychologically meaningful.

Topic:

Knowledge structures

Talks / 39

Efficient Computation of Stability Rates in Adaptive Assessments Based on Knowledge Structure Theory

Authors: Luca Stefanutti[1]; Debora de Chiusole[1]; Alice Jenisch[2]; Pasquale Anselmi[1]

[1] *University of Padua*

[2] *University of Tuebingen*

Computerized adaptive assessments require stopping rules that effectively balance efficiency and accuracy. Recently, the stability rate has been proposed as a novel stopping criterion for adaptive assessments grounded in knowledge structure theory. The stability rate is defined as the probability that the inferred knowledge state remains unchanged after administration of all remaining items. The assessment terminates once this probability exceeds a predefined threshold. A direct computation of the stability rate during an adaptive assessment requires generating all possible continuations (binary response vectors) compatible with the observed partial response pattern. Because the computational cost increases exponentially with the number of remaining items, this approach rapidly becomes infeasible in practice. To address this issue, we propose more efficient algorithms that evaluate only a subset of all possible continuations. The Best-to-Worst (B2W) algorithm computes the stability rate cumulatively, beginning with the most favorable continuation, whereas the Worst-to-Best (W2B) algorithm proceeds in the opposite direction, starting from the least favorable continuation. B2W is expected to be more efficient when the stability rate is low, while W2B should perform better when the stability rate is high. The performance of both algorithms, as well as several hybrid variants, is investigated through a series of simulation studies focusing on computational efficiency and accuracy.

Topic:

Knowledge structures

Poster / 51

Recovering Knowledge Structures from Incomplete Data with Inductive Item Tree Analysis

Authors: Peter Steiner[1]; Aliaksei Badnarchuk[1]; Debora De Chiusole[2]; Milan Segedinac[3]; Goran Savić[3]; Luca Stefanutti[2]; Jan Hochweber[4]

[1] *St. Gallen Universtiy of Teacher Education*

[2] *University of Padua*

[3] *University of Novi Sad*

[4] *University of Teacher Education St.Gallen*

Effective adaptive instruction depends on knowing what a student has mastered and what they are ready to learn next. Traditional scale scores, such as those obtained from item response (IRT) models, provide only limited support for this purpose. Consequently, several recent approaches conduct student assessment within hierarchical structures of test items or competencies. While these approaches are theoretically promising, practically applicable methods for identifying such structures from empirical data remain scarce, particularly in the presence of missing item responses.

To address this gap, this contribution introduces MAR-IITA—a new implementation of Inductive Item Tree Analysis (IITA) that provides three variants of handling missing responses. To obtain first insights into its performance, the following research question was addressed: At what percentage of missing responses can the proposed variants still recover the underlying structure?

To answer this question, response data were simulated along a structure under different noise conditions. Missing responses were introduced incrementally across conditions, and performance was assessed in terms of the proportion of missing data up to which the true structure could still be recovered. Additionally, the same procedure was applied to a subset of the PISA 2003 dataset based on a previously identified structure. For each dataset, the three variants were evaluated using 1,000 replications.

On average, the algorithm recovered the underlying structure with a missing response percentage of 28.5% (SD = 4.5pp) for the noise-free simulation, 19.9% (SD = 4.4pp) for the simulation with noise, and 15.4% (SD = 4.6pp) for the PISA data.

As this study examines a first implementation and is restricted to a limited combination of conditions and structures, the results provide an initial estimate of missing-data tolerance while highlighting substantial opportunities for further optimization.

Topic:

Knowledge structures

Poster / 50

Predicting Individual Distress in Exposure Therapy: An Explainable Machine Learning Approach

Authors: Annika Trapple[1]; Daniel Reiter[2]

Coauthors: Alexander Karner ; David Steyrl ; Dominik Pegler ; Filip Melinscak ; Frank Scharnowski ; Julian Weinhuber ; Mengfan Zhang

[1] *University Vienna*

[2] *University of Vienna*

Exposure therapy is considered the clinical gold standard for treating anxiety disorders. Yet its labor-intensive nature and reliance on trained therapists severely limit scalability. We propose supporting this process by developing a system to automate stimulus choice based on the successful prediction of fear-responses. The aim is to create a therapy process that is not only more efficient, but also highly personalised.
In order to implement this, the system must be able to reliably induce predefined levels of distress by selecting stimuli that elicit targeted fear responses. We employed an explainable machine learning pipeline using LightGBM to predict subjective units of distress (SUDs) based on demographics, clinical questionnaires and image-specific attributes. Utilizing SHAP values allowed insights into the relationship between model and data features. To optimize performance within clinical constraints, we used a custom cross-validation function to simulate information gain during the therapy process. Model accuracy was evaluated while progressively incorporating subsets of subject-specific SUDs from the testing into the training set. We compared models incorporating random individual SUDs with models using only the lowest SUDs per person. The results indicated that integrating samples of participants' fear-rating patterns (SUDs) into the training set strongly improved prediction accuracy, with performance scaling positively as the number of included ratings increased ($n \in \{0, 5, \cdots 40\}$). Mean explained variances (R^2) were significantly above chance levels while mean absolute errors (MAEs) were significantly below chance levels in all models. This enables a closed loop system, where models can be updated with each new reaction to a stimulus, thus enabling the exposure therapy process to be individualised while also improving efficiency.
Ongoing work involves integrating these models into a closed-loop framework that combines behavioural predictions with real-time physiological signals, such as heart rate and eye tracking, in order to iteratively calibrate stimulus selection.

Topic:

Computational methods

Talks / **46**

Interpretability as Validity Evidence: A Claim-Based Workflow for Machine-Learning Explanations in Psychological Measurement

Author: Stefan Coors

Coauthors: Bernd Bischl ; Clara Sophie Vetter [1]; Jonas Hauck ; Sven Hilbert

[1] *University of Munich, Munich Center for Machine Learning*

Supervised machine learning is increasingly used for prediction in psychological and educational research, but model explanations are often interpreted without specifying the inferential claim they are meant to support. This creates a validity problem: technically correct explanation outputs can be narrated as evidence about psychological mechanisms, interventions, or decision rules even when they support only weaker claims about the fitted model or the observed covariate distribution.

We propose a claim-based workflow for interpretable machine learning in tabular psychological data. The framework treats explanations as validity evidence for explicitly scoped interpretive claims. It distinguishes three explanation semantics: within-support descriptive summaries of model behavior, model-based what-if queries about the fitted prediction function, and causal or recourse claims that require additional assumptions and feasibility constraints. These semantics are linked to diagnostic gates that specify what evidence is needed before an interpretation can be reported as defensible.

The workflow is designed for psychological and educational data settings in which predictors and outcomes may be scale scores, factor scores, item responses, registry variables, or plausible values, and in which observations may be clustered, weighted, dependent, or measured with imperfect reliability. Core diagnostics address measurement readiness, out-of-sample predictive adequacy, support and dependence, interaction heterogeneity, calibration and decision utility, local-explanation faithfulness, stability, predictive multiplicity, transport, and subgroup comparability.

We illustrate the workflow using two applied examples: a psychological screening task and an educational achievement task. In both cases, the central object of evaluation is not the explanation method alone, but the interpretive claim: what is being claimed, under which semantics, for which population and measurement context, and with what evidence. The proposed workflow complements existing prediction-model reporting standards by making the evidential warrant for machine-learning explanations explicit, reviewable, and aligned with psychometric validity reasoning.

Topic:

Psychometrics

Poster / 45

How Cohort Composition Shapes Latent Space Geometry: Benchmarking Multimodal Signatures Using an mlr3 mbSPLS Framework

Author: Clara Sophie Vetter[1]

Coauthors: David Popovic ; Nicolaos Koutsouleris ; PRONIA Consortium

[1] *University of Munich, Munich Center for Machine Learning*

A fundamental challenge in computational psychiatry is determining how population recruitment and data modality selection alter the latent structures learned by multivariate models. To address this, we developed a scalable, open-source pipeline utilizing the mlr3 machine learning framework to implement multiblock sparse partial least squares (mbSPLS), a method explicitly designed to uncover shared latent components across high-dimensional, multimodal data blocks. This framework provides a standardized, reproducible workflow for multi-block data integration across diverse neuroscientific context by integrating nested cross-validation, permutation testing for component significance, and bootstrapping for latent variable stability assessment.

In this study, we leveraged the framework to benchmark how sample composition influences these discovered latent spaces. Using the multisite PRONIA cohort (239 baseline features across 9 clinical and neurocognitive domains), we estimated outcome-agnostic multimodal signatures under two distinct sampling strategies: one using patient data only, and another incorporating healthy controls. The resulting latent expressions were tested for predicting i) remission in recent-onset psychosis and ii) transition to psychosis in clinical high-risk individuals and recent-onset depression.

Both sampling designs successfully recovered an identical primary latent dimension linking psychosocial burden, protective factors, objective cognition, and functioning. However, secondary latent signatures were highly sensitive to sample choice, shifting from fine-grained, symptom-specific phenotypes in the patient-only design to broader health-to-disease gradients when controls were included. Consequently, prognostic utility was strictly outcome-specific: patient-only representations optimally predicted localized remission endpoints, whereas control-inclusive models were required to robustly predict clinical transition.

These findings demonstrate that study population selection explicitly dictates the geometry of learned latent spaces and subsequent clinical classification. Crucially, this work showcases the utility of the mlr3 mbSPLS framework as a versatile tool for multimodal latent variable discovery, with ongoing applications extending to other neuroscientific and longitudinal datasets to further evaluate the stability of psychiatric signatures across different clinical contexts.

Topic:

Computational methods

Talks / 40

Generalisation and confidence in decisions from experience

Author: Rebecca West[1]

Coauthor: Thorsten Pachur [1]

[1] *Technical University of Munich*

Real-world decisions often require people to evaluate unfamiliar risky options based on limited experience with similar situations. In this study, we investigate how people learn the statistical structure of risky environments and use this knowledge to evaluate novel options. Participants first learned how shape and colour mapped onto the mean and variance of payoff distributions by repeatedly observing outcomes from a set of exemplar stimuli. Participants then evaluated both the learned exemplars and novel stimuli that varied systematically in similarity to the learned options, providing both preference and confidence judgments. To characterise the mechanisms underlying this process, we combined models of risky choice with models of generalisation. These models were used to examine how people inferred reward statistics across the stimulus space, and how these beliefs shaped preference judgments and confidence. This approach allowed us to test whether generalisation differed across the two statistical dimensions of risky choice, as well as whether individual differences in risk sensitivity related to the way people learned and generalized information. Together, these findings provide insight into the computational principles underlying generalisation, risky choice, and metacognitive evaluation in novel decision environments.

Topic:

Mathematical models of psychological processes

Talks / 42

Novel Approaches for Testing Intertemporal Decision-Making Preferences of Individuals

Author: Sofia Wolfson[1]

Coauthors: Debora De Chiusole [1]; Luca Stefanutti [1]

[1] *University of Padua*

This study investigates delayed gratification and delay discounting by integrating theoretical frameworks from the fields of quantitative psychology, behavioral economics, and mathematical modeling. Delayed gratification refers to the ability to resist immediate rewards in favor of larger future outcomes, whereas in economics, delay discounting refers to the relative valuation of a good over time. While delay discounting and time preferences put a monetary value on the reward itself in different points in time, delayed gratification relates to the present cognitive process of evaluating the benefits and costs of accepting an immediate reward.

Discount delay is measured by two mathematical models characterized by exponential and hyperbolic equations and are used to describe the discounting curve and the area under the curve (AUC). AUC is used because its distribution is not skewed, and it does not require assumptions about the

mathematical form of the discounting function. Most models reduce delay discounting to a single parameter, the discounting rate, which has clear limitations; area based measures compare individuals only by willingness-to-pay, ignoring the curve's slope and the decay of gratification over time. These shortcomings call for new methods to interpret time-value tasks.

In this study we applied McFadden's Random Utility Theory (RUT) as an alternative framework to traditional hyperbolic discounting models, treating the hyperbolic model as a special case within the broader RUT approach introduced by McFadden. Using pairwise comparison data, the study applies additive conjoint measurement and compares RUT with Birnbaum's True and Error model for choice behavior. The analysis further examines differences across variables such as age, socioeconomic status, and financial background using data collected from over 250 participants. To the best of our knowledge, this represents a novel application of RUT to delayed gratification pairwise questions.

Topic:

Other (please comment above)

Talks / 38

Relative Distinctiveness, Interference, and the Revised Feature Model

Author: James Yearsley[1]

Coauthors: Dominic Guitard [2]; Jean Saint-Aubin [3]; Marie Poirier [1]

[1] *City St Georges, University of London*

[2] *Cardiff University*

[3] *University of Moncton*

The Revised Feature Model (RFM) is a memory model based on the principle that recall depends on the similarity between traces of items in primary memory and stored representations in longer term memory. Forgetting is attributed to similarity-based retroactive interference rather than passive trace decay, and rehearsal can, in some cases, counteract this interference.

Across a number of papers, the RFM has been applied to model immediate and delayed versions of serial recall, reconstruction, and free recall, both for word lists and for spatial locations. A particular success of the model has been in accounting for the 'production effect'–the effect where items spoken aloud, or otherwise produced when presented, are subsequently better recalled, but the model has also been applied to explain the impact of distractor tasks on recall, and the effect of increasing time between item presentation on recall, based on the effect on rehearsal.

In this talk, we will focus on how these effects emerge from the model, and in particular how the interaction between distinctiveness, interference, and rehearsal necessitates a formal model to make predictions. We will explain how the model has led to a number of empirical predictions, and evaluate how well they have been confirmed. We will also discuss what the model does not include, and how it demonstrates the importance of considering proper benchmark models when developing models to explain specific experimental effects.

Topic:

Mathematical models of psychological processes

Poster / 44

Developing tools to measure story-based reasoning

Author: James Yearsley

A huge literature in psychology, particularly in the qualitative tradition, stresses the importance of stories as way in which people reason about themselves, other people, and the world around them. In decision making, for example, some Authors: have argued that we construct stories that connect events in the past as a way of explaining the present and predicting the future.

However, relatively little is known about how to quantitatively measure the stories people believe. This makes it challenging to test theories, or to develop ways to use stories to promote behaviour change. We present work attempting to use causal structure as a framework for representing story-based reasoning. We developed tools to elicit judgments about causal relations between events, and compared these to endorsements of explanations, predictions, and counterfactual judgments.

Our work shows qualitative agreement between derived measures of story structure and counterfactual reasoning, but there are significant discrepancies. We discuss what this means for developing computational models of story-based reasoning.

Topic:

Measurement and scaling