Contents

Welcome

Dear Colleague,

it is our great pleasure to welcome you to the 2025 Meeting of the European Mathematical Psychology Group held at the University of Padua, Italy. We hope that, like the previous editions, even this 56th EMPG meeting will be stimulating for new future scientific work. At the same time we hope you will enjoy the city of Padua with its "historic flavor".

This year's conference features three keynote talks by Jürgen Heller, Dora Matzke, and Michel Regenwetter.

Finally, we would like to acknowledge the financial support by the FISPPA department and the Department of General Psychology.

The organizing committee:
Luca Stefanutti
Pasquale Anselmi
Debora de Chiusole
Andrea Spoto

General information

Conference organization

Scientific committee
Rocío Alcalá-Quintana (Complutense University of Madrid, Spain)
Hans Colonius (University of Oldenburg, Germany)
Clintin Davis-Stober (University of Missouri, USA)
Adele Diederich (University of Oldenburg, Germany)
Jean-Paul Doignon (Université Libre de Bruxelles, Belgium)
Ehtibar Dzhafarov (Professor Emeritus of Purdue University, USA)
Miguel Ángel García-Pérez (Complutense University of Madrid, Spain)
Matthias Gondan (University of Innsbruck, Austria)
Jürgen Heller (University of Tuebingen, Germany)
Luigi Lombardi (University of Trento, Italy)

Local organizing committee

Luca Stefanutti (Chair)
Pasquale Anselmi
Giovanni Bruno
Debora de Chiusole
Ottavia Epifania
Umberto Granziol
Federica Morleo
Stefano Noventa
Andrea Spoto
Sofia Wolfson

Conference web site and e-mail
Web site: `https://psychometrics-laboratory.psy.unipd.it/EMPG2025`
E-mail: empg2025.meeting@unipd.it

The venue

The conference takes place in an area called "Cittadella dello Studente" (colored area in the picture below), which is composed of 5 buildings. All sessions will take place in room **T3**, located inside Building 4, named "Centro Linguistico di Ateneo". Coffee breaks will take place in room **AMU**, located inside Building 5, named "Casa dello Studente".

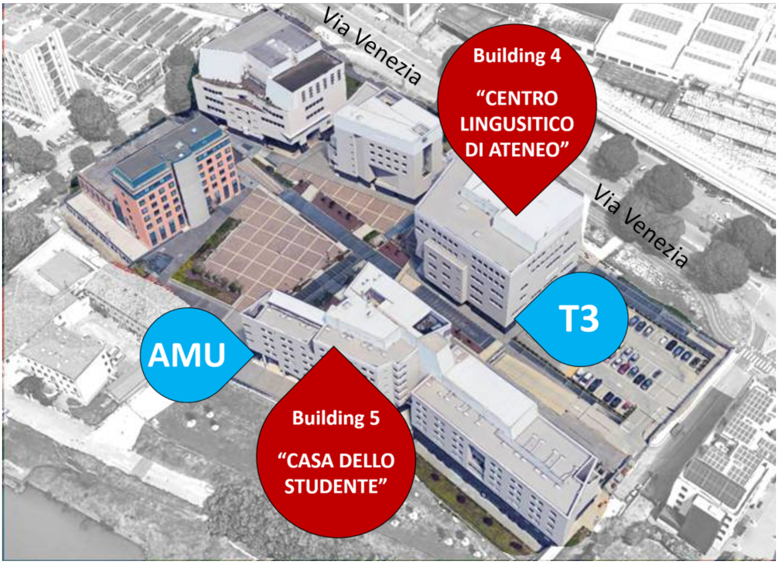

Figure 1: Map of the "Cittadella dello Studente" area.

Addresses

Centro Linguistico di Ateneo (Building 4)
University of Padua
Via Venezia, 16
35131, Padova (Italy)

Casa dello Studente (Building 5)
University of Padua
Via Venezia, 20
35131, Padova (Italy)

Registration desk

Participants can register for the conference, collect conference materials, and obtain all relevant information at the registration desk, located in room **AMU** inside Building 5, named "Casa dello Studente". During the conference days (September 3rd to 5th), the registration desk will be open from 8:30 a.m. to 6:00 p.m. On Tuesday, September 2nd, it will be open from 6:00 p.m. to 8:00 p.m.

Presentation guidelines

Talks

Talks are scheduled for 20 min, including 5 min for discussion. It is important that speakers and chairs strictly adhere to the time schedule to offer everyone the opportunity to switch back and forth between sessions. In all rooms you will find the option to connect your own computer or tablet to the presentation equipment. If you plan to use the local computer (Windows operating system), please hand your presentation in PDF or PPT format to the technical assistant present in the room well before the session starts.

Keynotes

This year, the EMPG conference will feature three keynote talks by Juergen Heller, Dora Matzke, and Michel Regenwetter. Each day of the three-day conference kicks off with their talks from 9:00 to 10:00 a.m.

Posters

The posters will be on display on September 4th, from 15:40 to 17.00 pm, during the Coffee break, and just outside and facing the entrance of room **AMU** (Building 5, named "Casa dello Studente"). Posters must fit within and be attached with Velcro, which will be provided, to the poster boards, that are 90 cm high (2.95 feet) by 70 cm wide (2.30 feet). Please attach your poster to the board before 2.30 pm.

Welcome session

The welcome session will take place September 2nd in room **AMU** (Building 5, named "Casa dello Studente") from 6.00 pm to 8.00 pm. There will be some refreshments and the opportunity to meet other conference participants.

Coffee breaks and lunch

Refreshments and snacks will be provided during coffee breaks on Tuesday, Wednesday and Thursday, in room **AMU** (Building 5, named "Casa dello Studente")

The area surrounding the departments and the School of psychology offers many different lunch options. Close to the Psychology departments there are a number of small snack bars serving pasta, sandwiches or salads. There are also a number of restaurants with typical Italian or international food. The following one is a list of suggested places:

Antico Desiderio (snack bar with pasta, salad and sandwiches)
Via del Portello, Padova
Phone: 049 807 7244

La Piadineria (student snack bar offering the so-called "piadina" a thin, soft Italian flatbread from the Emilia-Romagna region).
Via Venezia 9/A, Padova
Phone: +39 349 2952220

Pizzeria al Porteo (pizza and pasta)
Via del Portello, 35131 Padova
Phone: 049 772311

Ristorante Venezia (pasta, meat, and fish restaurant)
Via Venezia, 28, 35131 Padova
Phone: 049 807 4920

Tre Scalini (student snack bar with pasta, salads and sandwiches)
Via Venezia, 2, Padova

Phone: 049 807 0489
Web site: http://www.bartrescalini.it/

Verba Volant (student snack bar with pasta, salad and sandwiches)
Via Niccolò Tommaseo 114, Padova
Phone: 049 7811251

Zafferano (fish restaurant)
Via Niccolò Tommaseo, 67, 35131 Padova
Phone: 049 807 9389
Web site: http://www.zafferanopadova.it/

Conference dinner

The conference dinner will take place at the beautiful restaurant EXFORO, which features a panoramic terrace overlooking the famous Prato della Valle. The elegant setting and refined cuisine will offer a memorable experience for all participants.

Travel

Padua can be reached by plane via Venice or Treviso airports:

- Via Venice - Upon arrival at Marco Polo airport you can reach Padua by: (1) private shuttle bus (needs to be booked 24 hours before) or by (2) ATVO public bus to Venezia-Mestre train station and then by train to Padua;
- Via Treviso - Upon arrival at A. Canova airport you can reach Padua by private shuttle bus (needs to be booked 24 hours before) or by public bus (line 6 of ACTT company) to Treviso train station, and then by train to Padua.

Accommodation

BEST WESTERN PREMIER Hotel Galileo****
Via Venezia 30, 35131, Padua - Italy

Phone: +39 049 7702222
Fax: +39 049 7800762
E-mail: galileo.pd@bestwestern.it
Web site: www.hotelgalileopadova.it/en/wellness.aspx

SAVHOTEL MANTEGNA⋆⋆⋆⋆
Via Tommaseo 61 , 35131, Padua - Italy
Phone: +39 049 8494 111
E-mail: mantegna@savhotel.com Web site: https://www.savhotel.com/mantegna/

BEST WESTERN HOTEL BIRI⋆⋆⋆⋆
Via Grassi, 2, 35129 - Padova - Italy
Phone:: +39 049 8067700
Fax: +39 049 8067748
E-mail: biri.pd@bestwestern.it
Web site: http://www.hotelbiri.com/

HOTEL MAJESTIC TOSCANELLI⋆⋆⋆⋆
Via dell'Arco, 2, 35122, Padua – Italy
Phone: +39 049 663244
Fax: + 39 049 8760025
E-mail: majestic@toscanelli.com
Web site: www.toscanelli.com/index.php/en/

HOTEL RISTORANTE AL SANTO⋆⋆⋆⋆
Via del Santo, 147, 35123, Padua- Italy
Phone +39 049 8752131
Fax: +39 049 8788076
E-mail: alsanto@alsanto.it
Web site: www.alsanto.it/indexing.htm

ALBERGO VERDI⋆⋆⋆
Via Dondi dall'Orologio, 7, 35139 Padova - Italy
Phone: 049 8364163
Fax: 0498780175
info@albergoverdipadova.it
Web site: www.albergoverdipadova.it/inglese/indice.htm

BELLUDI37
Via Luca Belludi, 37, 35100 Padua - Italy
Phone: +39 049 665633
Fax: +39 049 658685
E-mail: info@belludi37.it
Web site: http://belludi37.one-lab.it/en/hotel/

HOTEL PLAZA
Corso Milano, 40, 35139 Padova - Italy
Phone: +39 049.656822
Fax: +39 049.661117
Email: info@plazapadova.it
Web site: http://www.plazapadova.it/en

Program

Wednesday, September 3

08.30 – 09.00	**Opening session**
09.00 – 10.00	**Keynote** A personal view on four decades of knowledge structure theory *Jürgen Heller*
10.00 – 11.00	**Knowledge space theory I** Chair: P. Anselmi State-dependent response error probabilities: A generalization of the BLIM *Alice Jenisch, Jürgen Heller* Reliability measures in knowledge structure theory *Debora de Chiusole, Andrea Spoto, Umberto Granziol, Luca Stefanutti* Extending the basic local independence model to multiple observed classifications *Pasquale Anselmi, Debora de Chiusole, Egidio Robusto, Luca Stefanutti*
11.00 – 11.30	**Coffee break**
11.30 – 12.30	**Perception & Psychophysics** Chair: D. Kohler

	Decoding the three operations of perceptual recognition *Jerald Balakrishnan*
	Comparing bisection-based and adaptive psychophysical methods for eliciting indifference points in preference research *Wei-Chen Chang, Yung-Fong Hsu*
	Investigating role independence of internal references in cross-modal magnitude productions *Dorina Kohler, Katharina Naumann, Jürgen Heller*
12.30 – 14.00	**Lunch break**
14.00 – 15.40	**Statistical methods** Chair: A. Spoto
	Imputation of missing outcome measures in meta-analysis *Irene Alfarone, Filippo Gambarota, Massimiliano Pastore*
	Does sampling matter in psychological science? *Alberto Arletti, Irene Alfarone, Livio Finos, Gianmarco Altoè*
	Detecting careless responding with the Resampled Response Reliability (ReReRe) measure *Vittorio Guerrieri, Marcello Passarelli*
	Post-Selection Inference in Multiverse Meta-Analysis (PIMMA): Enhancing validity and transparency in meta-analytic inference *Matteo Manente, Filippo Gambarota, Livio Finos, Anna Vesely, Gianmarco Altoè*
	Improving inter-rater reliability estimation in dichotomous classification with multiple raters *Andrea Spoto, Anna Panzeri, Umberto Granziol, Giovanni Bruno, Michele Vicovaro, Giuseppe Mignemi, Massimo Nucci*

15.40 – 16.10	**Coffee break**
16.10 – 17.10	**Psychometrics** Chair: O. M. Epifania The self calibrating score method to develop a questionnaire *Giovannibattista Flebus* Infinite dimensional psychology: Nonparametric methods in psychological sciences *Giuseppe Mignemi, Anna Panzeri, Giovanni Bruno, Andrea Spoto* Say "when": An item response theory algorithm for shortening tests while accounting for response fatigue *Ottavia M. Epifania, Livio Finos, Luigi Lombardi*

Thursday, September 4

09.00 – 10.00	**Keynote** Winner takes all: Flexible cognitive "race" architectures for response inhibition *Dora Matzke*
10.00 – 11.00	**Cognitive models** Chair: A. Heathcote Context independence assumption in the stop-signal paradigm: No systematic violations *Michelle Donzallaz, Henrik Godmann, Andrew Heathcote, Dora Matzke* Modelling visual working memory: A computational approach for continuous response space data *Alberto Petrin, Francesco Resta, Roberto Dell'Acqua* Suitability of evidence accumulation models for conflicting choices

	Marit Metz, Malte Luken, Stefan Radev, Andrew Heathcote
11.00 – 11.30	**Coffee break**
11.30 – 12.50	**Knowledge space theory II** Chair: S. Noventa Questioning the question: The impact of item selection rules in adaptive assessment *Andrea Brancaccio* Recent developments in McCLASS *Matthias Gondan, Jeremy Irilli, Irene Alfarone* Stability rate and a dynamic stopping criterion in KST-based adaptive assessment *Luca Stefanutti, Debora de Chiusole, Pasquale Anselmi* Toward a unified perspective on assessment models: KST, CDM, and IRT models with continuous latent variables *Stefano Noventa, Jürgen Heller, Tim F. Schaffland, Augustin Kelava*
12.50 – 14.00	**Lunch break**
14.00 – 15.40	**Decision making** Chair: J.-P. Doignon Comparing the 2N-ary Choice Tree model and the Cube model for best-worst choice situations with three choice alternatives *Adele Diederich* Discovering predictive and interpretable models of decision-making via hybrid neural networks and symbolic regression *Simone D'Ambrogio, Jan Grohn, Nima Khalighinejad, Marcelo Matter, Laurence Hunt, Matthew Rushworth* Extrapolating beyond exemplars: A new computational model of quantitative judgment

Seitz Florian I., Rieskamp Jörg, von Helversen Bettina

How to cast and evaluate moral theories as competing parsimonious models of choice
Michel Regenwetter, Brittney Currie, Yu Huang, Sylvia E, Anna K. Carlson

Around Falmagne's representation theorem for finite random scale systems
Jean-Paul Doignon

15.40 – 17.00 **Coffee break & Poster session**

A formal approach to psychoanalytic theory: A prelogical setting from a quantum model
Giulia Battilotti, Rosapia Lauro Grotto

IRT model selection dealing with psychological constructs: A conflict between theory and practice? Evidence from a Monte Carlo simulation.
Giovanni Bruno, Andrea Spoto

Potentials in Markov random fields and dependences in visual perception. Cues on plausible correspondences
Luigi Burigana

Extending deep belief networks with accumulation-based decision processes
Francesco Maria Calistroni, Antonino Visalli, Alberto Testolin, Marco Zorzi

Psychometric validation of the Cooperative and Competitive Personality Scale (CCPS) in an Italian sample
Stefano Dalla Bona, Marta Caserotti, Giovanni Bruno

Optimizing the assessment of mathematical skills through competence-based knowledge structure theory
Paul Doerrbecker, Jürgen Heller

Leveraging idiographic network analysis for precision interventions in artificial psychology: Modeling dynamic symptom interactions
Hojjatollah Farahani, Peter Watson, Natasa Kovač

Frontiers and neighborhoods of a clinical state
Umberto Granziol, Andrea Brancaccio, Andrea Spoto

From EPIC data to cognitive insights: Detecting decline with the shifted Wald model
Tuba Hato, Mischa von Krause, Andreas Voss

Evaluating soft skills in school teachers: An assessment grounded in CbKST
Federica Morleo, Alessandra Vitanza, Pasquale Anselmi

BayesFlow Neo: Fast and flexible amortized Bayesian inference for everyone
Stefan T. Radev, Lars Kühmichel, Valentin Pratz, Paul-Christian Bürkner

Bayesian discovery of cognitive dynamics: A Gaussian process diffusion decision model
Lukas Schumacher, Stefan T. Radev

Combining theory- and data-driven methods to build knowledge spaces: The case of linear functions
Peter Steiner, Jan Hochweber, Michael Kickmeier-Rust, Stephanie Leininger

A hitchhiker's guide to information theoretical measures in psychology
Niels Van Santen, Yves Rosseel, Daniele Marinazzo

The foreign language effect in the illusion of causality: A replication attempt based on Bayes factor design analysis
Michele Vicovaro, Eduardo Navarrete, Stefano Dalla Bona

17.00 – 18.00 Business meeting

Friday, September 5

09.00 – 10.00	**Keynote** “Wait, what are you saying, exactly?” A theoretical framework for codifying and evaluating verbal hypotheses about proportions *Michel Regenwetter*
10.00 – 11.00	**Measurement & Scaling** Chair: E. Dzhafarov Extended tetrad analysis in factor modelling: Separability and uncertainty from multidimensional dependence structures *Mario Angelelli* How consumers perceive the Nutri-Score scale *Rodion Iurev, Víctor Temprano-García, María de Haro González, José Luis García-Lapresta* Epistemic odds of contextuality *Ehtibar N. Dzhafarov, Janne V. Kujala, Víctor H. Cervantes*
11.00 – 11.30	**Coffee break**
11.30 – 12.30	**Bayesian methods** Chair: N. Stevenson Bayesian evidence synthesis and heterogeneity *B. Alfons Edmar, Duco Veen, Irene Klugkist* Bayesian experimental design with mutual information and learned errors for human-computer interaction *Hugo Miquel, Julien Gori, Olivier Rioul* Generalized Bayesian hierarchical structural equation modeling *Niek Stevenson, Andrew Heathcote, Dylan Molenaar, Birte Forstmann, Dora Matzke*
12.30 – 14.00	**Lunch break**

14.00 – 15.40	**Reaction times** Chair: H. Colonius Analyzing response time distributions in mental rotation tasks *Lukas Legner, Lisa Bechtum, Matthias Gondan* Testing the race model inequality using linear quantile mixed models *Luigi Lombardi, Lorenzo Arena* Projected spherical diffusion model: A tractable evidence accumulation theory of estimation *Amir Hosein Hadian Rasanan, Sebastian Olschewski, Jörg Rieskamp* A computational framework for integrating Predictive processes with evidence Accumulation Models (PAM) *Antonino Visalli, Francesco Maria Calistroni, Margherita Calderan, Francesco Donnarumma, Marco Zorzi, Ettore Ambrosini* Quantile function and mixture distribution revisited *Hans Colonius, Adele Diederich*
15.40 – 16.00	**Closing session**

Abstracts

Keynote Speakers

A personal view on four decades of knowledge structure theory

Jürgen Heller
University of Tuebingen, Department of Psychology, Germany

In 1985 Jean-Claude Falmagne and Jean-Paul Doignon introduced a new non-numerical approach to knowledge assessment, providing a detailed picture of the individual strengths and weaknesses in a particular domain, and forming the basis for personalized learning. This seminal work initiated a highly active research strand of research with an ever-growing literature, and has lead to successful large-scale applications. The initially deterministic theory implementing a behavioristic perspective has been elaborated and extended in various ways: introducing a probabilistic framework, accounting for underlying cognitive elements, generalizing the response format to polytomous items, clarifying the connection to related approaches, and widening the scope beyond educational applications – to name a few. In my talk I will highlight some of these contributions with a focus on the last decade (following up on Jean-Claude's keynote on thirty years of knowledge structure theory at the EMPG 2015), and I will reflect on possible future directions.

Winner takes all: Flexible cognitive "race" architectures for response inhibition

Dora Matzke
University of Amsterdam, The Netherlands

Imagine you're driving down the highway, stuck behind a slow-moving car. You glance in the rear-view mirror to check if it's safe to overtake, but just before doing so, you hear the siren of an approaching ambulance—and immediately abort the maneuver. This ability to stop an ongoing action when it becomes inappropriate is known as response inhibition. Response inhibition is a core component of executive control and is essential for navigating an ever-changing and often unpredictable world safely and effectively. Inhibitory ability is commonly measured by the stop-signal reaction time (SSRT), which reflects the time required to complete an inhibitory process after receiving a stop signal. Since SSRTs are not directly observable, they must be inferred using a cognitive model in which independent inhibitory ("stop") and response ("go") processes race with each other to control behavior. In this talk, I present a flexible cognitive architecture that extends the traditional race model, enabling researchers to study response inhibition across a wide variety of situations and paradigms relevant to cognitive control. Introduce different parameterizations of this framework, highlight their respective strengths and limitations, and demonstrate their utility using data from experimental, clinical, and neuropsychological investigations. I will discuss models that incorporate features of real-world stopping, such as attentional lapses, interactions between going and stopping, and the overlap of inhibitory processes involved in action stopping and interference control. I will conclude by exploring the potential of this framework to offer a comprehensive understanding of the mental processes that govern behavior in dynamic environments—especially in situations requiring complex decision sequences and adaptive control.

“Wait, what are you saying, exactly?” A Theoretical Framework for Codifying and Evaluating Verbal Hypotheses about Proportions

Michel Regenwetter
University of Illinois Urbana Champaign, USA

Verbal hypotheses about multiple proportions abound across behavioral science. I will provide a general-purpose theory-building framework, grounded in deductive science. Using the accompanying online app, scholars can codify a joint system of hypotheses into a single mathematical formal model, thereby disambiguating their verbal claims. No mathematics knowledge beyond high-school level is required. Instead of conceptualizing their scientific theory as a, potentially ill-specified, alternative to a ‘no-effect’ strawman Null hypothesis, scholars can compare multiple competing substantive claims theoretically and empirically. The proposed theoretical framework enables one to quantity the parsimony (hence refutability) of each hypothesis, characterize and quantify the overlap among competing claims, formally model heterogeneous populations, characterize the predictions of each hypothesis succinctly, and provide insights into the edge cases of each hypothesis. We also cast predictability and replicability as testable mathematical models. By helping the scholar steer clear of over-specifying, over-fitting, and/or over-predicting, the framework opens up new avenues of depth and nuance in the development and analysis of theories about multiple proportions.

Regular Talks & Posters

Imputation of missing outcome measures in meta-analysis

Irene Alfarone[a], Filippo Gambarota[b], Massimiliano Pastore[b]
[a] *University of Innsbruck,* [b] *Univerisity of Padova*

Meta-analyses in psychology often pool outcomes measured using different instruments, with many studies reporting only a subset of these outcomes. Such instruments might measure the same underlying construct and be correlated, making multivariate meta-analysis (MVMA) a methodologically appropriate choice. However, missing outcomes raise concerns about bias and inference validity, particularly when the missingness mechanism is unknown. While some outcomes may be Missing Completely at Random or Missing at Random, a more complex scenario arises when outcomes are Missing Not At Random. We propose a strategy for addressing missing study-level summary statistics using repeated stochastic single imputation. Within each Monte Carlo replicate, missing effect estimates, standard errors, and within-study correlations are replaced by a single random draw from predefined donor distributions (e.g., uniform, univariate normal, multivariate normal), to capture a range of plausible data-generating mechanisms. By doing so, we systematically explore the interaction between assumptions about the missingness mechanism and alternative meta-analytic specifications across thousands of replicates. The proposed approach has two main advantages: it is computationally efficient and, most importantly, makes the assumptions underpinning the analysis explicit rather than relying solely on the data. This transparency allows researchers to clearly articulate and test the implications of their assumptions for a straightforward implementation in sensitivity analyses of psychological MVMA.

Extended tetrad analysis in factor modelling: Separability and uncertainty from multidimensional dependence structures

Mario Angelelli
University of Salento (IT)

Geometric representations offer a principled framework for describing latent constructs and clarifying sources of uncertainty in their dimensional characterisation. This talk introduces a novel geometric formulation of factor models, based on subspaces spanned by a pair of matrices, one encoding quantitative information such as factor scores or estimates, the other encoding structural information. Within this framework, determinantal expressions explicitly quantify the contribution of subsets of dimensions to the overall factor structure. This approach refines rank-based conditions central to understanding factor score indeterminacy and highlights the implications of non-uniqueness in instrumental variable estimation in over-identified settings. By assigning weights to these multidimensional contributions, we can analyse sensitivity to their variation consistent with determinantal constraints. This scaling enables an algebraic procedure extending the definition of tetrads to our setting, establishing conditions for identifying variability components attributable to each individual dimension. We derive minimal conditions—formulated in terms of graph planarity—that ensure identifiability of dimension-specific contributions. The methods developed to prove these results also provide formal verification tools and the construction of counterexamples where identifiability fails. These counterexamples reveal a novel form of uncertainty in factor models, termed contextuality. Specifically, when the weighted contributions to the association between the two factor subspaces are ranked by magnitude, we show that the pairwise comparison between any two components depends on the choice of additional elements used to complete a basis with them. This dependence on "irrelevant alternatives" violates rational preference principles. We conclude by relating this form of contextuality to specific notions of uncertainty discussed in the psychometric literature.

Extending the basic local independence model to multiple observed classifications

Pasquale Anselmi, Debora de Chiusole, Egidio Robusto, Luca Stefanutti
Univerisity of Padova

Do the probabilities of mastering language items differ between individuals with and without autism spectrum disorder? Do the probabilities of careless responding to arithmetic items differ between individuals with and without dyscalculia? Does gender affect these probabilities? The talk presents the multiple observed classification local independence model (MOCLIM), a probabilistic model specifically developed to address these and similar questions. Developed within knowledge structure theory, the MOCLIM extends the basic local independence model by allowing the probabilities of knowledge states (each being the set of items an individual is capable of solving) and the careless error and lucky guess probabilities of items to differ among populations. Individuals are classified into specific proficiency, careless, and guessing classes, each defined based on one or more observed individual variables (e.g., gender, education level, medical diagnosis). All individuals within the same proficiency, careless, and guessing classes share the same knowledge state, careless error, and lucky guess probabilities, respectively. MOCLIM parameters can be estimated via maximum-likelihood using the expectation-maximization algorithm. Results from a simulation study suggest that the true parameter values are well recovered by the estimation algorithm and that the true model can be identified by comparing the goodness-of-fit of alternative models. An empirical application to data from Raven-like matrices indicates that the MOCLIM effectively discriminates between situations where group differences are expected and those where they are not.

Does sampling matter in psychological science?

Alberto Arletti[a], Irene Alfarone[b], Livio Finos[c], Gianmarco Altoè[c]
[a]*Ca' Foscari University of Venice*, [b]*Universität Innsbruck*, [c]*University of Padova*

Researchers in Psychological Science frequently rely on convenience samples (e.g., college students) in their studies. This practice rests on the tenet that the characteristics of many psychological constructs (e.g., personality traits) are universal and shared across all human beings. As a result, there are often the tacit assumptions that 1) how individuals are sampled in a study require little attention and 2) that larger samples are more representative. However, low-quality sampling can lead to selection bias, which in turn can potentially invalidate the statistical analysis. We demonstrate how careless sampling can derail both population estimates (e.g., how many highly neurotic over 65 are there in Italy) and coefficient estimation (e.g., what is the relationship between neuroticism and age). Moreover, we show that increasing sample size without addressing bias can paradoxically make the situation worse. We conclude by offering researchers theoretical tools to identify and prevent selection bias, alongside practical statistical and methodological countermeasures.

Decoding the three operations of perceptual recognition

Jerry Balakrishnan
California Polytechnic University

Perceptual recognition judgments are the outcomes of three interconnected operations: encoding, response selection, and response execution. In this paper, I show that the states and interdependence of these three processes are mirrored in the dynamics and interdependence of three outcome probabilities, each conditioned on the state of the system at a given time: the stimulus, the accuracy of the eventual response, and the response. By estimating these values from single-cell neural activity, therefore, it is possible to establish which of the three operations are being carried out in a given region of the brain.

A formal approach to psychoanalytic theory: A pre-logical setting from a quantum model

Giulia Battilotti, Rosapia Lauro Grotto
University of Florence

Our goal is the reading of the features of the structural Unconscious in a formal logical setting. We build on abstract views proposed in the development of psychoanalytic theory, on the basis of the original Freudian theory of representation, of Matte Blanco's approach to the logic of the

Unconscious via Bi-Logic, and of the abstract approach to psychoanalysis, proposed by Klein and Bion. Our work is set on a logical representation of quantum states that leads to the idea of infinite singleton. Such an idea meets the requirements of Matte Blanco's infinite sets and characterize the symmetric mode of Bi-logic, since it allows to see the differentiation between the Freudian thing presentation with respect to the word presentation, in formal terms. Then we interpret the components of thinking, corresponding to the symbiotic, convivial and parasitic links characterized by Bion, by a modal operator , and discuss the role of the infinite vs finite elements for the mind in relation to time. As a consequence one can see the transition from the Primary to the Secondary process in formal logical terms. Moreover, one can see that the reading of the spin states in the Bloch sphere can characterize the two primitive abstract opposite attitudes implemented in our mind, namely the separating and the integrating one.

Questioning the question: The impact of item selection rules in adaptive assessment

Andrea Brancaccio
National Research Council of Italy

In Knowledge Space Theory, adaptive assessment procedures typically rely on three key components: a questioning rule, an updating rule, and a stopping rule. The most commonly used questioning rule is the half-split rule, which selects the next problem such that the probability of the test-taker mastering it is equal to the probability of failing it. However, this approach does not account for item-specific parameters such as careless error and lucky guess probabilities, nor does it take into consideration the specifics of the updating rule used in the procedure. This presentation introduces and analyzes alternative questioning rules that explicitly incorporate these factors. We explore specific scenarios in which these new rules may offer advantages — for example, in the presence of equally informative items or large disparities between item parameters (e.g., $\beta_q \gg \eta_q$). Both theoretical and simulation-based results are presented to evaluate the accuracy and efficiency of the proposed questioning strategies compared to the standard approach.

IRT model selection dealing with psychological constructs: A conflict between theory and practice? Evidence from a Monte Carlo simulation.

Giovanni Bruno, Andrea Spoto
University of Padova

When testing the psychometric properties of a psychological assessment tool, selecting the most appropriate IRT model depends on several factors, including model fit and the theoretical assumptions that guided the development of the scale. For example, we may assume that all items on our new psychological scale are equally capable of differentiating individuals with varying levels of a latent trait (e.g., ability). This assumption would imply fixing the discrimination parameter across items. Similarly, we might assume that the psychological distance between response categories is consistent across items, designing them to share the same scale structure (e.g., a 5-point Likert scale); this would mean fixing the item-category thresholds. These assumptions would lead us to select a specific IRT model—such as the Rating Scale Model (RSM)—whose characteristics align with our theoretical framework. However, this alignment is not always maintained in practice. Model selection is often driven by model fit comparisons alone or by arbitrary a-priori decisions. Such practices tend to favor more flexible models (e.g., the Partial Credit Model [PCM] or the Generalized Partial Credit Model [GPCM]), which, while providing a better empirical fit, may contradict the theoretical assumptions underlying the scale's development. As a result, we risk describing the scale in a way that disregards its theoretical foundations, estimating parameters that were initially intended to be constrained. This creates a non-trivial conflict between the theoretical aims that informed scale construction and its actual implementation. Such divergence may ultimately pose challenges for both contextual interpretation and research applications. This contribution starts from the premise that when the properties of an IRT model and the test's theoretical assumptions are well aligned, the most restricted IRT model may be the most appropriate choice—not in spite of its simplicity, but because it was designed to fit such data. Although more flexible models may consistently outperform it in terms of fit indices, they may do so at the cost of theoretical coherence.

Potentials in Markov random fields and dependences in visual perception. Cues on plausible correspondences

Luigi Burigana
University of Padova

So-called "potentials", defined as functions on small sets of topologically neighboring variables, are the building blocks of Markov random fields. These fields have gained a conspicuous position in the theory of computer vision, and recently have received consideration also in the psychological research on human visual perception. In my poster contribution, I present some cues concerning plausible conceptual similarities between potentials, as meant in the theory of Markov random fields, and dependences involved in perceptual field dynamics, both of the psychophysical and of the intra-perceptual kinds. To take account of such similarities may be useful in promoting formal modeling in perceptual psychology using probabilistic graphical models, as Markov random fields are, in which topological aspects, represented by graphs, are combined with probabilistic components, as factors of probabilistic inference in the formation of visual scenes. For illustration, I make reference to salient perceptual phenomena, such as perceptual grouping, phenomenal transparency, and lightness perception.

Extending deep belief networks with accumulation-based decision processes

Francesco M. Calistroni[a], Antonino Visalli[b], Alberto Testolin[a], Marco Zorzi[a]
[a] *University of Padova,* [b] *University of Modena and Reggio Emilia*

Deep Belief Networks (DBNs) are stochastic recurrent neural networks that learn a hierarchical generative model of sensory data. Downstream perceptual and cognitive tasks are typically implemented as a read-out from the internal representations (latent states) through a classifier that produces dichotomous outputs. While effective for modeling accuracy data in psychophysical tasks, these outputs do not capture the dynamics of decision-making observed in biological agents. Human and animal behavior, particularly in perceptual decision-making, is better described by evidence accumulation models—such as leaky integration, drift diffusion, or race models—which provide a temporally extended account of how decisions emerge over time from noisy sensory evidence. In this work, we integrate evidence accumulation processes with DBN-based representations. Specifically, we use the internal representations of trained DBNs as input features to feed into classical evidence accumulation models. This approach allows us to bridge the gap between high-dimensional, learned neural representations and psychologically grounded decision dynamics. This hybrid model can be used to simulate both the accuracy and temporal dynamics of decision behavior, capturing response time distributions and speed-accuracy trade-offs, which standard DBNs classifiers cannot replicate. Embedding biologically inspired decision mechanisms into deep learning architectures is a crucial step toward more cognitively plausible models of artificial perception and action. This integration enables not only a better alignment with behavioral data but also offers insights into how hierarchical feature extraction in neural networks can be mapped onto sequential decision-making processes.

Quantile function and mixture distribution revisited

Hans Colonius, Adele Diederich
University of Oldenburg

Two statistical concepts play important roles in cognitive modeling, specifically in the analysis of reaction times and response probabilities: (a) the quantile function and (b) the mixture of probability distributions. Examples abound: *q-q* plots, delta plots, and vincentization for (a); dual-process models, fast-guess models, two-state learning models, etc. for (b). Here we explore ways to bring these two concepts together and present new results, e.g., on quantile mixtures. Non-parametric results are illustrated by parametric distribution examples, like Weibull, Gumbel, and other distributions.

The foreign language effect in the illusion of causality: A replication attempt based on Bayes Factor Design Analysis

Michele Vicovaro, Eduardo Navarrete, Stefano Dalla Bona
University of Padova

The illusion of causality is a cognitive bias for which individuals mistakenly infer a causal relationship between two unrelated events. This bias has been studied in associative learning contexts, where people observe varying co-occurrences of a potential cause and its supposed effect. A normative index used to assess causality is the $\Delta P = P(\text{Effect}|\text{Cause}) - P(\text{Effect}|\text{not Cause})$. In some illusory conditions – such as when the effect occurs frequently regardless of the presence of the cause (e.g., $P(\text{Effect}|\text{Cause}) = P(\text{Effect}|\text{Not Cause}) = 0.75$) – ΔP equals zero, indicating no causal relationship. Nevertheless, participants commonly overestimate the strength of the causal connection, providing ratings significantly above zero on a 0 to 100 Likert scale. Diaz-Lago and Matute (2019) found that this illusion can be reduced when the task is conducted in a foreign language – a phenomenon known as the Foreign Language Effect (FLE). They proposed that the cognitive disfluency associated with processing information in a non-native language promote more analytical thinking, leading to more normative (i.e., accurate) causal judgments. However, recent findings by Dalla Bona and Vicovaro (2024) challenge this interpretation, highlighting the need for further research on the mechanisms behind the FLE. Our study aimed to replicate Diaz-Lago and Matute's findings by comparing two groups of participants completing the task underlying the illusion – one in Italian (native language, NL) and one in English (foreign language, FL). A secondary objective was to explore potential cognitive mechanisms contributing to the FLE in causal reasoning. To determine the necessary sample size for detecting the two group differences, we employed a simulation-based approach informed by distributional characteristics of causal judgment data reported by Dalla Bona and Vicovaro (2024). These simulations allowed us to construct a prior distribution for Cohen's d, which we used within a Bayes Factor Design Analysis (BFDA) framework to determine sample size (110 participants for each group). The results of our study did not replicate (BF01 = 365 for the null model) the original findings by Díaz-Lago and Matute (2019). This outcome underscores the critical role of replication in psychological science and emphasizes the need for studies with sufficient statistical power.

Psychometric validation of the Cooperative and Competitive Personality Scale in an Italian sample

Stefano Dalla Bona, Marta Caserotti, Giovanni Bruno
University of Padova

In the analysis of human behavior within social contexts, a fundamental conceptual distinction is frequently drawn between cooperativeness and competitiveness. This dichotomy frames scenarios wherein individuals must navigate the tension between contributing to the collective welfare and promoting personal interests. Although the legitimacy of this binary classification has been subject to extensive debate, it remains a valuable framework for modeling behavior and has recently garnered attention in the domain of personality research. Within this conceptual framework, the Cooperative and Competitive Personality Scale (CCPS) was developed to reliably assess and distinguish between these two behavioral tendencies. The present study aimed to evaluate the psychometric properties of the scale within an Italian sample (N = 701). While the original validation study proposed a two-factor model, positing cooperation and competition as independent constructs, an Exploratory Factor Analysis (EFA; N = 352) conducted in the current study revealed a hierarchical structure. Specifically, the analysis indicated the presence of two negatively correlated second-order factors (i.e., cooperativeness and competitiveness) suggesting that these dimensions may potentially represent opposing poles on a single continuum rather than single constructs. Each second-order factor is composed of three first-order factors, representing the construct (i.e., cooperativeness or competitiveness) articulated across three domains: (i) behavioral, (ii) cognitive, and (iii) affective. Subsequently, a Confirmatory Factor Analysis (CFA; N = 349) employing the Diagonally Weighted Least Squares (DWLS) estimation method was conducted using R. The results provided satisfactory support for the proposed model, as indicated by the fit indices (RMSEA = 0.02, CFI = 0.99, TLI = 0.99). Furthermore, both subscales demonstrated adequate internal consistency, with McDonald's Omega coefficients exceeding the commonly accepted threshold of 0.80. Our validation process optimized the scale by reducing the number of items, thereby creating a more practical assessment tool. At the same time, we provided evidence that competitiveness and cooperation can be conceptualized as two opposing poles of the same construct, contributing to the ongoing discussion regarding the unidimensionality or duality of competitiveness and cooperativeness.

Discovering predictive and interpretable models of decision-making via hybrid neural networks and symbolic regression

Simone D'Ambrogio[a], Jan Grohn[a], Nima Khalighinejad[a], Marcelo Matter[b], Laurence Hunt[a], Matthew Rushworth[a]
[a] *University of Oxford*, [b] *New York University*

Computational cognitive models provide valuable insights into the cognitive mechanisms underlying a particular behavior. While this framework is valued for its interpretability, it is often limited in capturing realistic biological processes. Conversely, deep learning approaches offer greater flexibility but typically lack interpretability. We introduce a computational framework integrating artificial neural networks (ANN) with cognitive modeling to derive interpretable models of decision-making. Our hybrid approach embeds ANN within a theory-based cognitive model. Subsequently, symbolic regression is applied to distill the ANN's learned computations into human-readable mathematical equations. This hybrid model demonstrated superior predictive accuracy for behavioral choices compared to simpler benchmarks via cross-validation and generalized successfully to an independent dataset. Furthermore, the model's derived variables significantly outperformed benchmarks in predicting BOLD activity, which was measured using ultra-high field (7T) fMRI. This workflow enables data-driven discovery and validation of interpretable, predictive models in cognitive science.

Reliability measures in knowledge structure theory

Debora de Chiusole, Andrea Spoto, Umberto Granziol, Luca Stefanutti
University of Padua

Within the framework of Knowledge Structure Theory (KST), this study introduces two novel indices to assess the reliability of knowledge state estimation: the expected accuracy rate and the expected discrepancy. The expected accuracy rate, defined as $\rho = P(K = \widehat{K})$, quantifies the probability that the estimated modal state $\widehat{K}$ matches an individual's true knowledge state K. The expected discrepancy, instead, measures the expected distance $E(D)$ between the estimated and true states, capturing the magnitude of error in cases of misclassification. We derive both indices analytically and examine their asymptotic properties, complemented by simulation studies to evaluate their behavior in finite samples. The simulation results highlight a trade-off between the number of items administered and the size of the knowledge structure: smaller structures maintain stable accuracy across varying error rates, while larger structures are more sensitive to errors. However, increasing the number of items leads to improved estimation accuracy even in complex structures. Importantly, the expected discrepancy remains low, suggesting that misclassified states tend to lie close to the true ones. An empirical application using real assessment data further confirms the practical utility of these indices, demonstrating that KST-based assessments can provide robust and interpretable diagnostic information. Overall, our findings support the use of KST as a reliable and informative framework for cognitive diagnosis in educational and psychological assessment.

Comparing the 2N-ary Choice Tree model and the Cube model for best-worst choicesituations with three choice alternatives

Adele Diederich
University of Oldenburg

The 2N-ary Choice Tree (2NCT) model (Wollschlaeger and Diederich, 2012, 2020) and the Cube model (Mallahi-Karai and Diederich, 2019, 2023) are dynamic-stochastic approaches for decision making situations with multiple alternatives. The 2NCT model is a dynamic stochastic model formalized as a random walk on a tree. It shares several features of other stochastic models on decision making such as including initial biases for any of the choice alternatives, updating preferences over time or initiation of a response when a decision criterion is met. Its distinct assumption is that it establishes two counters for each alternative for tracking evidences in favor of choosing a specific alternative and a negative counter for tracking evidence against choosing that alternatve. It provides a mechanism to predict a preference order of the the offered alternative The multi-episode Cube model postulates that best–worst choice task is the outcome of sequential choices made in a number of episodes allowing the alternatives to be ranked from best to worst or from worst to best. The underlying model is a multivariate Wiener process with drift issued from a point in the unit cube, where episodes are defined in terms of a sequence of stopping times. Both models make predictions with respect to choice probabilities and (mean) choice response times. It is shown how the models can be implemented using Markov chains and how they are are tested on data from (Hawkins et al., 2014b).

Optimizing the assesment of mathematical skills through competence-based knowledge structure theory

Paul Dörrbecker, Jürgen Heller
University of Tuebingen

Competence-based test optimization is a method to improve the efficiency of an existing test by creating a new version of the test which contains the minimal number of Items while extracting the maximal amount of information from the data. This method is based on the competency deletion procedure (CDP) developed by Anselmi et al. (2022) and an approach of only using one of multiple equally informative items based on the sum of error probabilities estimated within a basic local independence model (BLIM; Falmagne and Doignon, 1988). The aim of this method is to make an existing test as informative as possible about the competence state (the set of skills an individual has available) of the participant while also reducing the number of items needed to asses this competence state as much as possible. The presented work applies this method to one of the most commonly used German elementary school numeracy tests, a speed test called “Heidelberger Rechentest 1-4” (HRT 1-4, Haffner et al., 2005), in particular to the subtest “Subtraction”. Research about elementary school level subtraction skills resulted in 8 different skills which the HRT 1-4 should be able to asses. Competence structures and skill functions were established based on an expert survey with five elementary school teachers. Applying CDP allowed for creating much shorter versions of the test, which are as informative as the initial version, but only need 5 to 7 items instead of the 40 items in the original version. This remarkable shortening opens up the possibility that students could solve the items without time pressure. Theoretically, this would lead to the same information on the students' competence states, without having to deal with the disadvantages of a speed test. Results on evaluating the empirical validity of this conclusion are reported. Data from more than 1000 students (first to fifth grade) were reanalyzed in order to investigate how well the shorter versions of the test perform compared to the original version in terms of uncovering the underlying competence state.

Around Falmagne's representation theorem for finite random scale systems

Jean-Paul Doignon
University of Bruxelles

A random choice function on some universe of alternatives specifies the probability $p(a, B)$ that the subject, when offered some set B of alternatives, selects the alternative a in the set. A classical model, the Random Scale Model (as in Block and Marschak, 1960), predicts the choice probabilities from a probability distribution on all linear orderings of the universe: $p(a, B)$ equals the probability that the subject's latent ordering puts a before all other elements of B. The Bradley-Terry-Luce model is another classical model, which predicts choice probabilities from latent utilities of alternatives: each alternative a in the universe has a utility $u(a)$, and the choice probability $p(a, B)$ is equal to the normalized value of $u(a)$. When no utility vanishes, Luce characterizes the random choice functions predicted by the BTL model in a straightforward way. To the contrary, describing the RSM predictions happens to be more delicate. For finite universes, Falmagne (1978) remarkably shows that the predictions are characterized by the nonnegativity of the so-called Block-Marschak polynomials (better, linear forms). For general universes, McFadden and Richter (1970) have another characterization (which, in the finite case, remains more involved than Falmagne's one). Of course, assessing the relevance of the two models requires experimental work. On the other hand, theoretical investigations can prove useful in model comparisons. For instance, in joint work with Davide Carpentiere, we establish that the RSM resists probabilistic resolutions, while the BTL model does not. A resolution (Cantone, Giarlotta and Watson, 2021) is a way of building a new choice function from some given ones (the base and the fibers). We easily extend resolutions to the probabilistic setting. In the second part of the talk, weak orders replace linear orders in the RSM. Choice correspondences then take the place of choice functions: the value $p(A, B)$ records the probability that the subject, when offered some set B of alternatives, selects the subset A of B. With Kota Saito, we build a characterization of the model predictions. The proof is similar to the illuminating proof of Falmagne Theorem by Fiorini (2004), in that it also relies on a well-chosen flow polytope.

Context independence assumption in the Stop-Signal Paradigm: No systematic violations

Michelle Donzallaz, Henrik Godmann, Andrew Heathcote, Dora Matzke
University of Amsterdam

Response inhibition, the ability to stop ongoing responses when they are no longer appropriate, is commonly assessed using the stop-signal paradigm. The paradigm owes its popularity to the underlying horse-race model that enables estimation of the unobservable latency of stopping. The model conceptualizes performance as a race between a go and a stop process. A key assumption of the model is that the go process is unaffected by the presence or absence of a stop signal. Bissett et al. (2021) found that this "context independence" assumption is commonly violated, especially when the stop signal is presented shortly after the primary task. As violations of context independence invalidate nonparametric estimates of stopping latencies derived from the race model, they should not be taken lightly. In this talk, we present a systematic assessment of context independence through an extensive review of the literature. Our review uncovered 2090 stop-signal papers published between 2000 to 2021, resulting in 31 papers and 107 conditions with usable and openly available datasets included in our analysis. We found no systematic violations of context independence. In fact, our results generally align with the predictions of a version of the race model that accounts for attention lapses, formalized as failures to trigger the stop process. We discuss the implications of these findings for applied researchers as well as the broader stop-signal literature.

Epistemic odds of contextuality

Ehtibar N. Dzhafarov[a], Janne V. Kujala[b], Víctor H. Cervantes[c]
[a]*Purdue University,* [b]*University of Turku,* [c]*University of Illinois Urbana-Champaign*

Any experiment can be presented as a system of random variables in which every random variable is double-indexed: by its content, which is the question it answers, and its context, the circumstances under which it is recorded. Such a system can be contextual or noncontextual, the latter term generalizing the notion of a variable being selectively influenced by its content only. The essence of contextuality is in the similarity of random variables having the same content in different contexts. To measure how similar two such variables are we compute the maximal probability with which they could coincide if they were jointly distributed (de facto no two variables are if they are in different contexts). A system is contextual if these same-content variables, in order to be compatible with other variables in their respective contexts, have to be more dissimilar than they are when considered in isolation, outside their contexts. Among systems of random variables a historically prominent role was assigned to cyclic ones. A cyclic system of rank n consists of $2n$ variables with n contents arranged into n contexts, each context containing precisely two variables and each content being shared by precisely two variables. What is the epistemic (Bayesian) probability that a randomly chosen cyclic system of rank n is contextual? As it turns out this probability rapidly tends to zero as n increases. Consequences of this finding are discussed.

Bayesian Evidence Synthesis and heterogneity

B. Alfons Edmar, Duco Veen, Irene Klugkist
University of Utrecht

In recent years there as been a resurgence in the interest in synthesising evidence in terms of test statistics as an alternative and compliment to the well established estimation framework. One of these methods, which we refer to as Bayesian Evidence Synthesis (BES), is particularly interesting due to it's simplicity and capability of combining sources of evidence that have substantial methodological diversity. This is because the aggregation of evidence occurs at the level of the hypothesis rather than the effect size. While this method can handle a large degree of methodological diversity, how it behaves when statistical heterogeneity is present in the sample is currently unknown. The current simulation study examines how BES performs under varying levels of statistical heterogeneity in a simple mean difference model. The results show that BES performs well in both low- and high-power scenarios, but substantial heterogeneity in small effects often makes for irreconcilable results in the BES framework, given that the hypothesis being tested is that each study included in the set of studies share the same directional hypothesis. We discuss the merits and potential pitfalls when employing BES and what type of model comparisons are needed for safe and reliable inference.

Say "when": An item response theory algorithm for shortening tests while accounting for response fatigue

Ottavia M. Epifania[a], Livio Finos[b], Luigi Lombardi[a]
[a]*University of Trento,* [b]*University of Padua*

Item Response Theory (IRT) offers a robust framework for shortening existing tests, as it provides detailed information into the measurement precision of each item across different levels of the latent trait. This makes it possible to design brief, yet psychometrically sound, short test forms

(STFs) that help mitigate quality loss in responses due to respondent fatigue. This study explores the estimation of latent trait levels under three conditions: (i) using all items from a full-length test without accounting for response fatigue, (ii) using all items while accounting for response fatigue, and (iii) using a subset of items selected by the Léon algorithm, which is designed to build informative STFs while factoring in fatigue effects. In this context, response fatigue is modeled by lowering the upper asymptote of the probability of a correct response (i.e., careless error). Simulation results show that ignoring response fatigue when administering all items leads to biased latent trait estimates, particularly at medium to high levels. This bias can be mitigated by administering a carefully selected subset of items that maintains the measurement precision of the full-length test.

Leveraging idiographic network analysis for precision interventions in Artificial Psychology: Modeling dynamic symptom interactions

Hojjatollah Farahani[a], Peter Watson[b], Natasa Kovač[c]
[a]*Tarbiat Modares University,* [b]*University of Cambridge,* [c]*University of Donja Gorica*

This paper explores a groundbreaking approach in Artificial Psychology, bringing together research on idiographic network analysis to transform how we personalize psychological treatment care. Imagine treatments designed just for you, easing struggles like anxiety or repetitive thoughts with pinpoint accuracy. Our synthesis shows how this innovative method creates highly personalized care by mapping each person's unique symptom patterns, meeting the urgent need for tailored solutions across psychological treatment challenges. We dive into studies that: (1) track symptoms in real time using smartphone-based surveys, (2) build connection maps showing how symptoms like anxiety and low mood interact at the same time, and (3) create prediction maps revealing how one symptom triggers another over time. These individual maps guide treatments targeting key symptoms, with ongoing data helping adjust care for better results. Using R code with example data, we illustrate how to build these maps, making the science practical and accessible. Focused on conditions like depression, our review highlights better symptom relief and patient involvement compared to standard care, while noting challenges like ensuring consistent results. This work paves the way for smarter, compassionate psychological treatment solutions, blending technology and human care.

The self calibrating score method to develop a questionnaire

Giovanni Battista Flebus
University of Milano-Bicocca

Contrary to the Likert scale method, the self-calibrating scoring method (A.k.a optimal score) allows for the development of scales based on contrasts between emotions, situations, images, and various alternatives. It takes the form of a multiple-choice question, as shown below: Item11 If I think about the current situation (Covid lockdown), I feel...() Bored () Abandoned () Angry () Perplexed () Helpless and vulnerable () Grateful for the health of my loved ones () Confident in the good resolution of the health emergency. Each option is assigned a weight in terms of z-scores, which are directly interpretable. The term “self-calibrating” refers to the method used to calculate the numeric score for each option and is the same as the scoring derived from correspondence analysis or the optimal score method. However, the name ‘self-calibrating score’ emphasizes that only the first dimension is considered crucial for interpretation. When applied to ordinal items like Likert scales, this method yields weights that marginally increase the alpha coefficient. Despite its early development, this method has rarely been utilized. However, this method becomes essential with multiple-choice questions, which can effectively portray diverse contrasts of emotions, situations, graphic images, and various other items. It can be used to measure attitudes, just like Likert scales, i.e. Self-assessed noise sensitivity, or attitudes towards assessment in school children. But the best results are obtained in measuring a complex or reasoning attitudes, where multiple psychological postures, attitudes, feelings, evaluations, self- descriptions, and self-assessments are compared to yield a total score of reasoning attitude, such as towards university exams or school life. To illustrate the power and potential applicability of this method, three published examples will be provided.

Recent developments in McCLASS

Matthias Gondan, Jeremy Irilli, Irene Alfarone
University of Innsbruck

We present recent developments of our E-learning platform McCLASS (Misconception-aware Competence Learning and Assessment Smart System), including separate R and Prolog packages that have spin off during development. These include automatic rendering of mathematical expressions in HTML, LaTeX and graphical representations in SVG, and a large set of functions for interval arithmetic to allow for rounding and imprecision in intermediate steps of a calculation. We present a method to classify the relevance of different steps in a calculation, and whether feedback should be given to specific mistakes or not. The system is extremely modular, and individual pieces can easily be integrated into other E-learning systems. We also give an outlook to its successor, MISTAKES (Mistakes-aware Intelligent Statistics Tutor for Assessing Knowledge and Enhancing Skills), and outline the planned efficacy evaluation.

Frontiers and neighborhoods of a clinical state

Umberto Granziol[a], Andrea Brancaccio[b], Andrea Spoto[a]
[a]*University of Padova,* [b]*National Research Council of Italy*

Computerized adaptive testing algorithms represent a promising advancement over conventional fixed-length questionnaires in clinical evaluations. Their capacity to select the most relevant items in real time, based on individual responses, leads to significantly shorter assessment durations while maintaining accurate estimations of psychological traits. Consequently, these adaptive tools serve as personalized short forms, finely tuned to the respondent's characteristics, without compromising psychometric integrity. In this paper, we propose that such tools may also offer insight into potential changes in clinical conditions. Within the framework of Formal Psychological Assessment, we introduce a methodological enhancement that incorporates the concepts of “frontier” and “neighborhood” derived from adaptive assessment outputs. The frontier includes symptoms that may soon develop or resolve as a disorder progresses, whereas the neighborhood refers to nearby clinical states that an individual might realistically transition into. We applied this approach to adaptive versions of two instruments designed to measure depression and generalized anxiety, using data from both clinical and non-clinical samples. We conducted simulations of adaptive assessments and evaluated whether they could replicate the same clinical classifications, neighborhoods, and frontiers as those identified through complete item responses, using comparable likelihood estimates. The simulated results mirrored those of full-length assessments while requiring fewer items due to the adaptive format. Additionally, the method allowed for reliable identification of both frontier and neighborhood components, yielding a compact yet informative clinical snapshot. This methodology demonstrates how adaptive assessments can extend beyond efficiency, offering clinicians data-driven, early-stage insights into possible disorder developments. As a result, they support quicker and more targeted therapeutic strategies, enhancing the clinical decision-making process.

Detecting careless responding with the resampled response reliability measure

Vittorio Guerrieri[a], Marcello Passarelli[b]
[a]*Italian Institute of Technology,* [b]*National Research Council of Italy*

Careless responding inflates measurement error and distorts model fit. However, identifying careless respondents is not a trivial task. Suggested methods include the even-odd index, which correlates factor halves, and extensions such as the resampled personal reliability. However, these methods falter when considering factors with a low number of items, and fail to flag as careless participants with low response variability (e.g., participants that answer “5” for most items). To overcome these limitations, we introduce the Resampled Response Reliability (ReReRe) measure. ReReRe operates by comparing an individual’s response consistency within theoretically related item sets (factor-based consistency) against their own baseline response consistency across randomly selected item sets. The rationale is that we expect that for careful participants, correlations that respect factorial structure would belong to the highest percentiles of potential correlation distributions. For each respondent, 1000 resamplings are drawn calculating: (1) correlation between two random halves of the entire survey, pairing items belonging to each factor; (2) correlation between two random halves of the entire item set without pairing. ReReRe is the upper-tail probability that the average resampled correlation that reflects the factorial structure belongs to that participants' distribution of inter-item correlations. Respondents with ReReRe $< .95$ (one-tailed) were considered careless. ReReRe on a dataset with 24 factors and $N = 468$ flagged 14% of respondents as careless, consistently with literature. We propose ReReRe as a distribution-agnostic advance on existing careless-response detectors, capable of identifying both randomly answering respondents and those exhibiting unnaturally uniform response patterns.

Projected spherical diffusion model: A tractable evidence accumulation theory of Estimation

Amir Hosein Hadian Rasanan, Sebastian Olschewski, Jörg Rieskamp
University of Basel

In everyday life, individuals frequently face situations that require estimating a continuous quantity and responding along a continuous scale. Examples range from simple tasks, such as estimating the time needed to walk to the train station, to complex financial judgments, such as forecasting the return on a specific stock investment. While evidence accumulation models are among the most successful frameworks for explaining choice behavior and response times, they have traditionally been limited to discrete choice tasks. Recent works have extended these models to accommodate decisions involving continuous response scales. However, a notable limitation of the current literature is the absence of mathematically tractable models for estimation tasks involving non-circular response scales—such as lines or arcs—which are commonly used in perceptual and cognitive estimation tasks. To address this gap, we build on recent developments in continuous evidence accumulation models to propose a comprehensive computational theory of estimation. Specifically, we extend the spherical diffusion model by projecting the accumulation process into a semi-circular space, resulting in a tractable diffusion model suitable for non-circular continuous scales. We derive analytical predictions for both estimation responses and response times. In particular, a central prediction of the model is that response distributions become increasingly skewed as stimuli approach the boundaries of the response range. To evaluate the model, we conducted two numerosity estimation experiments and fit the model to these datasets, as well as to three existing datasets: two involving numerosity estimation and another involving line length estimation. The empirical results were consistent with the model's predictions, demonstrating its ability to account for observed behavioral data. Overall, the proposed model generalizes to a broad class of tasks requiring responses within a closed continuous range—such as confidence ratings and probability judgments—and offers new insights into the cognitive mechanisms underlying human estimation and valuation.

Suitability of evidence accumulation models for conflicting choices

Marit Metz[a], Malte Luken[a], Stefan Radev[b], Andrew Heathcote[a]
[a]*University of Amsterdam*, [b]*Rensselaer Polytechnic Institute*

Eisenberg et al. (2019) reported a data set in which a largest number of participants (500) performed seven difference choice tasks requiring self-regulation due to the presence of conflicting information about the correct response. Each individual performed a relatively modest number of trials (100-200) on each task, and a subset returned for a second test session (see Enkavi et al., 2019, for an analysis of measures of the degradation of performance caused by conflict). We investigated the suitability of a variety of different types of evidence-accumulation models (i.e., cognitive models of decision processing) to this data on two grounds: 1) model comparison (i.e. which type of model provided the best account of each task in terms of being descriptively accurate and parsimonious), 2) parameter estimation (i.e., which type of model provided the best account of each task in terms of the precision, reliability and interpretability of parameter estimates). We discuss how our results can be used to refine efforts to more reliably identify and measure common psychological constructs underpinning self-regulation (Weigard et al., 2021, Stevenson et al., in press).

Comparing bisection-based and adaptive psychophysical methods for eliciting indifference points in preference research

Wei-Chen Chang, Yung-Fong Hsu
National Taiwan University

The bisection method has been implemented within the so-called tradeoff paradigm (Abdellaoui, 2000; Wakker & Deneffe, 1996) to elicit indifference points for utility assessment in decision theory. However, its precision may be compromised by two main factors: response randomness in participants' choices and the boundary determination rules inherent in the bisection algorithm. This study aims to evaluate the validity of bisection-based methods – including a simplified version – and to explore adaptive psychophysical alternatives such as ASA, PEST, and MOBS for eliciting indifference points. Two simulation studies were conducted to compare the performance of these methods under the experimental design of Abdellaoui et al. (2016) for measuring the value function. The simulation results show that, under a common stopping criterion, all methods are largely unbiased. However, the bisection method can produce biased estimates if boundary settings are poorly chosen. Among the tested methods, ASA demonstrates the highest efficiency, though it typically requires more iterations to complete. When the number of iterations is limited (but not too small), both the standard bisection method (with fixed initial boundaries) and the simplified bisection method perform efficiently for participants exhibiting more deterministic choice behavior. For participants with greater choice randomness, ASA and the simplified bisection method prove more effective. Given the prevalence of participant heterogeneity and practical constraints on iteration counts in real experiments, the simplified bisection method emerges as a reasonable compromise. We illustrate the application of this method in an experiment assessing loss aversion.

How consumers perceive the Nutri-Score scale

Rodion Iurev, Víctor Temprano-García, María de Haro González, José Luis García-Lapresta
University of Valladolid

Nutri-Score is a widely adopted front-of-pack nutritional labeling system that evaluates key nutritional components, including energy, saturated fat, sugars, and sodium, while positively scoring beneficial elements such as fiber, protein, and fruit/vegetable content. These parameters are synthesized into a five-level, color-coded ordinal scale ranging from A (green) to E (red). This labeling serves as a quick reference for consumers to rank product quality, facilitating informed purchasing decisions within a category and encouraging food manufacturers to reformulate products towards healthier options. Ordinal scales, such as Nutri-Score, are often assumed to be uniform, implying equal psychological distances between adjacent terms. However, this assumption does not always hold in practice. To address this, García-Lapresta and Pérez-Román (Appl. Soft Comput., 2015) introduced the notion of ordinal proximity measure, which assign an ordinal degree of proximity to each pair or terms in an ordinal scale, adhering to a specified set of axiomatic conditions. Later, García-Lapresta, González del Pozo and Pérez-Román (Inf. Sci., 2018) introduce the notion of metrizable ordinal proximity measure to enable metric-based interpretations of such ordinal relationships. Building on this theoretical foundation, García-Lapresta, Iurev and Pérez-Román (Expert Syst. Appl., 2025) developed a practical slider-based tool that allows respondents to visually indicate the perceived proximity between ordinal categories. The collected data are processed using purpose-designed software to construct metrizable ordinal proximity measures that reflect individuals' subjective perceptions. In this contribution, we apply this methodological approach to examine how consumers perceive the five-term Nutri-Score scale.

State-dependent response error probabilities: A generalization of the BLIM

Alice Jenisch, Jürgen Heller
University of Tübingen

The basic local independence model (BLIM) is one of the most prominent probabilistic models in knowledge structure theory. Its assumptions include the constancy of item-specific response errors across participants. So guessing the correct answer to an unknown item is assumed to be equally likely for a student who is completely new to the topic and a student with prior knowledge who is ready to learn the item. An analogous and equally implausible assumption holds for careless errors. Instead it would be desirable to let an item's error probabilities depend on how far an item is from the person's knowledge state. Different options for such a discrepancy are presented and a generalized local independence model with state-dependent error probabilities is introduced. First results of simulations using this generalized model are shown.

Investigating role independence of internal references in cross-modal magnitude productions

Dorina Kohler, Katharina Naumann, Jürgen Heller
Unversity of Tübingen

In cross-modal magnitude production experiments, the perceived intensities of stimuli are compared across different modalities, e.g. adjusting the brightness of a light to match the loudness of a sound or vice versa. We investigated psychological mechanisms underlying cross-modal intensity judgments by examining two well-known phenomena: firstly, the regression effect, where participants tend to shorten the range of whichever stimulus they adjust. Secondly, cross-modal commutativity, where the order of successive magnitude productions affects the final adjustments. Building on theoretical results from the global psychophysical model (Heller, 2021, 2025), we investigated whether these phenomena are related to role-dependent internal references. That is, reference points on the perceived intensity scale that depend on whether the corresponding stimulus serves as standard or target. In the experiment, visual and auditory stimuli are presented with standard intensities in a broad range of luminances and sound pressure levels. Additionally, three different production factors were used (1×, 2×, and 3×), as well as basic and successive trials. The variety of conditions improves parameter estimation. Using a Bayesian approach,

we test two main hypotheses: (1) A model with role-dependent internal references provides a better fit to the magnitude production data at the individual level, compared to assuming role-independence. (2) Especially for subjects showing violations of cross-modal commutativity, the model with role-dependent internal references provides a better fit.

Analyzing response time distributions in mental rotation tasks

Lukas Legner, Lisa Bechtum, Matthias Gondan
University of Innsbruck

Systems Factorial Technology is a non-parametric method by Townsend and Nozawa (1995) to analyze response time distributions in factorial experiments. Their method can distinguish between different mental architectures and stopping rules of the decision process. We have extended their method to account for both correct and errored response times in difficult tasks, that is, tasks with less than perfect accuracy. Here we demonstrate the application of this method to three mental rotation tasks. One task is a classical two-factorial design for a same-different task on two-dimensional polygon shapes, applying both rotation and distortion as experimental manipulations. The second task is a 4-level one-way setup in which the two experimental manipulations are projected on single dimension. Our results are consistent with serial exhaustive processing of these spatial transformations. In the third task, we rotate three-dimensional Shepard-Metzler figures around two cardinal axes. Our results show parallel exhaustive processing for some participants, which is interpreted as a single rotation around a skewed axis. We argue that adopting systems factorial technology in the analysis of cognitive experiments allows for deeper insights into the details of cognitive processes – retaining the rich structure of response time distributions and going beyond mere averages and elimination of errored responses. We also discuss limitations of the method due to model mimicry.

Testing the race model inequality using linear quantile mixed models

Luigi Lombardi, Lorenzo Arena
University of Trento

The race model inequality (RMI), originally proposed by Miller (Cognitive Psychology, 14, 247–279, 1982), defines an upper limit on the statistical facilitation of reaction times (RTs) that can be achieved by a race model within the redundant-signals paradigm (RSP). An observed violation of the RMI may serve as empirical evidence favoring a coactivation model over a race model. In this paper, we illustrate a semi-parametric statistical approach for assessing RMI at the group level, which also accommodates experimental manipulations in mixed-design studies. The method relies on linear quantile mixed models (Geraci and Bottai, 2014) and, notably, does not assume a specific distributional form for the RT variables involved in the RMI. Specifically, we formulate the RMI evaluation as a conditional quantile regression model for RTs, incorporating subject-specific random intercepts to account for within-subject dependence inherent to the RSP. Results from two Monte Carlo simulation studies demonstrate that the proposed method maintains type I error control while offering satisfactory statistical power. We conclude by presenting an empirical application to multisensory data.

Post-selection inference in multiverse meta-analysis: enhancing validity and transparency in meta-analytic inference

Matteo Manente[a], Filippo Gambarota[a], Livio Finos[a], Anna Vesely[b], Gianmarco Altoè[a]
[a]*Univerisity of Padova (IT)*, [b]*University of Bologna (IT)*

The replication crisis in psychology stems partly from researchers' degrees of freedom in data collection, coding, and analysis - often referred to as "the garden of forking paths". This flexibility creates a "multiverse" of equally valid but possibly divergent outcomes, where only favorable results are typically reported. Multiverse Analysis (MA) addresses this by exploring all plausible data and model choices, revealing the robustness of effects.

However, to date MA is purely descriptive and lacks inferential power. To overcome this, Post-selection Inference in Multiverse Analysis (PIMA) was introduced. PIMA uses a multivariate permutation framework to control for Type I error while maintaining statistical power. It applies to Generalized Linear Models (GLMs), enabling researchers to perform valid statistical inference across the multiverse. A potential application of PIMA is in meta-analysis, where results could be affected by arbitrary decisions - such as study selection, missing data handling, and model choice. By adapting PIMA to Multiverse Meta-Analysis (PIMMA), researchers can gain a transparent and statistically rigorous inferential framework for evaluating meta-analytic findings. We will present the effectiveness of this approach through simulation studies and real-world applications supported by ad hoc R functions. We will then conclude by providing practical recommendations for implementing the proposed procedure.

Infinite dimensional psychology: Nonparametric methods in psychological sciences

Giuseppe Mignemi[a], Anna Panzeri[b], Giovanni Bruno[b], Andrea Spoto[b]
[a]*Bocconi University*, [b]*Univerisity of Padova*

Discrete Bayesian nonparametric (BNP) priors have received increasing attention in recent decades. They characterise a broad class of flexible models in which inference is drawn under minimal distributional assumption and latent clusters are naturally accommodated by the structure of these priors. This makes the BNP models an appealing statistical solution for modeling individuals' heterogeneity in psychological sciences, in which overdispersion of participants' variability in real data poses serious concerns. Moreover, the verification of distributional assumptions under the parametric models might not be straightforward and the estimates might be severely affected by model misspecification. The BNP models overcome these issues as they estimate the parameters' density distribution from the data under less strict assumptions. Despite the flexibility of these methods, their applications in psychological sciences have been poorly investigated. Their theoretical complexity and the difficult implementation of the algorithms for the posterior computation have made these priors less considered than other Bayesian solutions within the most common models for social sciences, such as multilevel models and Item Response Theory (IRT) models. Significant theoretical and methodological progress has recently been made, and applying these methods is now more straightforward. We propose some real case examples in which discrete BNP priors are a convenient solution for exploring latent individual similarities and preventing inferences from overdispersion and model misspecification issues. Limitations and future directions are discussed, and user-friendly tutorial codes are provided to enhance a more direct and practical use of these methods in applied social and psychological sciences.

Evaluating soft skills in school teachers: An assessment grounded in CbKST

Federica Morleo[a], Alessandra Vitanza[b], Pasquale Anselmi[a]
[a] *University of Padova,* [b] *National Research Council of Italy*[b]

The Collaborative and Teamwork Skill test (CoTeSt) is an innovative assessment tool developed, validated, and applied in the framework of Competence-based Knowledge Space Theory (CbKST). It aims to accurately evaluate collaborative and teamwork skills in pre-primary and primary school teachers. The CoTeSt is a computer-based test that immerses participants in realistic team interactions using conversational agents within a narrative framework. It consists of a total of 20 sequential items in both multiple-choice and short-answer formats, which require six independent skills to be solved: (a) using proactive communication strategies, (b) using active listening techniques, (c) recognizing type and relevance of conflict situations, (d) identifying conflict resolution strategies, (e) planning action strategies, (f) monitoring performance and providing feedback. Each item is linked with one or more skills assumed to be relevant for its resolution, forming the competence model. The test is fully informative, that is, it allows for uniquely uncovering the set of skills a teacher masters (called teacher's *competence state*) based on the collection of items the teacher is capable of solving (called teacher's *knowledge state*). This work presents the identification of the teacher's competence state. A total of 161 primary and pre-primary school teachers completed the CoTeSt, and responses were coded as correct or incorrect before the analysis. The assessment of the teachers' competence states from item responses was conducted through a two-step procedure. First, the teacher's knowledge state was inferred based on their item responses. A likelihood was calculated for each knowledge state, reflecting the plausibility of it being the teacher's knowledge state. The knowledge state with the highest likelihood was selected as the teacher's uncovered knowledge state. Subsequently, the teacher's competence state was derived from their knowledge state, based on the competence model. The median of the likelihood of the teachers' uncovered knowledge states is .773, suggesting quite a high level of confidence that the uncovered knowledge states reflect the teachers' actual knowledge states. Based on the results of the assessment, the 161 teachers are characterized by 18 different competence states (out of the 64 possible). Overall, the results suggest that the CoTeSt is a useful tool for assessing teachers' collaborative and teamwork skills.

Toward a unified perspective on assessment models: KST, CDM, and IRT models with continuous latent variables

Stefano Noventa[a], Jürgen Heller[b], Tim F. Schaffland[b], Augustin Kelava[b]
[a]*University of Padua,* [b]*University of Tübingen*

In the past years, several theories for assessment have been developed within the fields of Psychometrics and Mathematical Psychology. The most notable are Item Response Theory (IRT), Cognitive Diagnostic Assessment (CDA), and Knowledge Structure Theory (KST). In spite of their common goals, these theories have been developed largely independently, focusing on slightly different aspects. Recently, Noventa, Heller, and Kelava (2024) proposed a general framework that allows to derive the models of these theories and systematize them within a general taxonomy by means of two primitives (structure and process) and two operations (factorization and reparametrization). In Noventa, Heller, Ye, Kelava (accepted) the relation between KST and CDA models was discussed as an application of the framework to the case of dichotomous latent variables. In this third contribution, IRT, CDA, and KST models are derived and discussed based on continuous latent variables. Implications for IRT models are discussed. A special focus is on the implicit representational assumptions required to move from the discrete case to the continuous one.

Modelling visual working memory: a computational approach for continuous response space data

Alberto Petrin, Francesco Resta, Roberto Dell'Acqua
Univerisity of Padova

Visual working memory (VWM) is a core component of the human cognitive architecture, and the nature of its limited capacity remains a highly debated topic in neuroscience. Over the past two decades, three conceptual frameworks have been developed to explain VWM capacity limits: (1) fixed-slot accounts (Zhang & Luck, 2008), (2) continuous resource accounts (van den Berg et al., 2014), and (3) interference-based accounts (Oberauer & Lin, 2017). Each of these accounts is formalized in a computational model that we fitted to data from 51 participants who performed a continuous report VWM task, in which a probed color had to be reported by selecting it on a color wheel. Critically, unlike most previous studies, we controlled for hue similarity by accounting for the degree of separation of the colors on the color wheel. We then performed a formal model comparison using the Akaike Information Criterion (AIC), also assessing each model's predictions with respect to two key aspects of participants' responses: the decline in precision with increasing set size and the occurrence of non-target color reports at higher set sizes (Bays et al., 2009). Our results showed that a model based on the idea of continuous resource allocation provided the best fit to the data, while those based on fixed-slot accounts, still considered the dominant framework, received the least support. Notably, a computational model implementing interference - based accounts - specifically developed to explain non-target color reports - did not perform optimally. In our study, these models failed to demonstrate superior accuracy in predicting non-target reports, suggesting either a suboptimal mathematical formulation or a mischaracterization of the underlying mechanisms.

BayesFlow neo: Fast and flexible amortized bayesian inference for everyone

Stefan T. Radev[a], Lars Kühmichel[b], Valentin Pratz[c], Paul-Christian Bürkner[d]

[a]*Rensselaer Polytechnic Institute,* [b]*Technische Universität Dortmund,* [c]*University of Heidelberg,* [d]*Universityof Dortmund*

Modern Bayesian inference involves a mixture of computational techniques for estimating, validating, and drawing conclusions from probabilistic models as part of principled workflows for computational modeling. This poster introduces the new and completely revamped BayesFlow software for amortized Bayesian inference (ABI). Along with direct posterior and likelihood estimation, the framework now includes multi-backend support for PyTorch, TensorFlow, and JAX, the fastest generative networks for sampling available, complete customization and high-level interfaces, as well as new capabilities for hyperparameter optimization, design optimization, and hierarchical modeling. A selection of cognitive benchmarks is presented.

How to cast and evaluate moral theories as competing parsimonious models of choice

Michel Regenwetter, Brittney Currie, Yu Huang, Sylvia E, Anna K. Carlson

University of Illinois at Urbana-Champaign

Philosophers and psychologists broadly agree that humans share universal moral values, but scholars disagree on the details. We analyze two leading theories in moral psychology, Social Domain Theory (SDT) and Moral Foundations Theory (MFT), in novel ways. We embrace contemporary calls for hardening psychological research that have emerged from the replication crisis. 1. We apply a Popperian philosophy of science ideal of formulating and testing parsimonious falsifiable theories. To that end, we represent SDT, MFT, as well as three sub-theories of MFT, through

restrictive mathematical constraints. 2. We avoid over- or mis-specifying hypothetical constructs by treating moral preference as an ordinal scale and probability as an absolute scale. 3. We translate the mathematical characterizations into formally precise, paradigm cross-cutting, probabilistic choice models. 4. Rather than follow "statistical rituals," such as hunting for effects or leaning on statistical models poorly grounded in substantive theory, we employ custom-designed order-constrained models and data analytics. We evaluate the empirical performance on several different levels of analysis, through both model-fitting and quantitative model-competition. Our findings reveal rich, detailed, and nuanced insights into SDT and MFT as either individual-level or universal theories.

Bayesian experimental design with mutual information and learned errors for human-computer interaction

Hugo Miquel[a], Julien Gori[b], Olivier Rioul[a]
[a] *Télécom Paris,* [a] *Institut des Systèmes Intelligents et de Robotique*

This work provides a Bayesian framework for handling user errors in interactive systems, with applications in human-computer interaction (HCI) and user modeling. The Bayesian Information Gain (BIG) algorithm is an iterative variant of Bayesian experimental design with mutual information as a cost function, used in HCI. It is a principled approach that maximizes expected information gained from each interaction. More precisely, let Θ be the potential target in the user's mind with prior distribution $p(\theta)$, X be the system feedback, and Y be the corresponding user's input. In each interaction loop, BIG selects feedback x that maximizes mutual information $I(\Theta; Y|X = x)$, assuming a known user model (likelihood) $p(y|x, \theta)$, and then updates the posterior distribution $p(\theta|x, y)$. This work extends the BIG algorithm to learn from user errors while preserving its mathematical foundations. We incorporate an error rate parameter ϵ into the likelihood function $p(y|x, \theta, \epsilon)$ and develop an adaptive algorithm that jointly infers both θ and ϵ by updating the posterior $p(\theta, \epsilon|x, y)$ at each interaction step. We also discuss three simplifying hypotheses for the prior expression $p(\theta, \epsilon)$ and three user models: (i) zero error; (ii) fixed error rate;

(iii) arbitrary random error rate. We prove mathematical continuity between these three models, showing that our adaptive approach naturally extends BIG. We also investigate model mismatch on the overall performance and degradation properties with respect to the standard BIG algorithm. While standard BIG converges quickly with perfect responses, it degrades with even small error rates. The fixed-error model depends critically on correctly estimating the error parameter, while our adaptive model achieves the highest accuracy under varying error conditions, at the expense of additional interactions.

Bayesian discovery of cognitive dynamics: A gaussian process diffusion decision model

Lukas Schumacher[a], Stefan T. Radev[b]
[a]*University of Basel,* [b]*Rensselaer Polytechnic Institute*

Cognitive modeling has a long tradition in attempting to capture the dynamics of mental processes. However, even though nearly everyone agrees on the importance of trial-by-trial changes in cognitive states, they are often oversimplified or entirely ignored in contemporary models. Thus, we introduce a novel model family called Gaussian Process Diffusion Decision Model (GPDDM). In his model family, key parameters of the Diffusion Decision Model are treated as evolving across trials, governed by mixtures of Gaussian process priors. To model a wide range of neurocognitively plausible trajectories, we combine multiple kernel functions: linear (e.g., learning or fatigue effects), periodic (e.g., rhythmic fluctuations), and radial basis function (e.g., smooth drift). This kernel composition affords both flexibility and interpretability in capturing practically relevant cognitive dynamics. To estimate GPDDMs, we employ amortized Bayesian inference, allowing for efficient and scalable posterior estimation. We demonstrate the validity of GPDDM in silico, showing that it can accurately recover latent parameter trajectories from data. Our approach offers a principled and flexible way to uncover cognitive dynamics from response data and opens new avenues for understanding how mental processes evolve over time.

Extrapolating beyond exemplars: A new computational model of quantitative judgment

Florian I. Seitz[a], Jörg Rieskamp[a], Bettina von Helversen[b]
[a]*University of Basel,* [b]*University of Bremen*

Quantitative judgments involve estimating numerical criterion values for objects based on their cues. These judgments are commonly explained by two cognitive model frameworks: rule-based models, which rely on learned cue-criterion relationships (like linear regression), and exemplar-based models, which rely on similar prior instances called exemplars. A key distinction between these models is their capacity to extrapolate: Rule-based models naturally extend predictions beyond trained ranges, while traditional exemplar models are limited to interpolation, computing a similarity-weighted average of the exemplars' criterion values. However, this distinction may be more methodological than theoretical. In this work, we introduce a novel exemplar model supporting extrapolation. Our model predicts an object's criterion value such that the object's differences from the exemplars in terms of both cues and criterion align as closely as possible. Specifically, the model assigns higher probability to criterion values for which the resulting criterion value differences to the exemplars deviate less from the fixed cue value differences. As a result, the model can assign high probabilities to criterion values outside the exemplar range when the cue values themselves are extrapolative. We demonstrate that the model produces approximately linear extrapolations across a variety of functional relationships, including non-linear and piecewise-linear functions—closely mirroring human extrapolation behavior documented in prior research. We further validate the model against both traditional exemplar and rule-based models using seven datasets spanning artificial and real-world domains. Our model consistently exhibits strong predictive performance, outperforming alternatives both on the aggregate and individual levels. On the aggregate level, the average evidence, measured with Akaike weights, exceeded the evidence for the traditional exemplar model by a factor of 2.05 and for the rule model by a factor of 1.36. On the individual level, 64% of participants were best described by our model. Overall, our findings reveal that extrapolation can emerge from a purely exemplar-based mechanism, offering a new perspective on how humans generalize from expe-

rience. This work contributes a formal, cognitively plausible account of exemplar-based extrapolation, with implications for theories of judgment and decision-making.

Improving inter-rater reliability estimation in dichotomous classification with multiple raters

Andrea Spoto[a], Anna Panzeri[a], Umberto Granziol[a], Giovanni Bruno[a], Michele Vicovaro[a], Giuseppe Mignemi[b], Massimo Nucci[a]
[a] *Univerisity of Padova,* [b] *Bocconi Univesity*

The reliability of psychological measures is crucial to ensure the validity and reproducibility of research findings. However, achieving inter-rater agreement in classification tasks involving subjective interpretation remains challenging, and traditional indices of inter-rater reliability often fail to account for the subjective biases influencing rater judgments. Two individual characteristics of raters are particularly relevant for improving inter-rater reliability estimates: the Belonging Measure (BM) and the Belonging Threshold (BT). BM reflects the rater's implicit belief that an item belongs to a category, expressed as a real number between 0 and 1. BT represents the minimum BM value at which the rater decides to assign the item to the category. Current indices—such as Krippendorff's α, Fleiss' κ, and Conger's κ—rely solely on observed classifications and disregard both BM and BT. This contribution proposes adjusted versions of these indices for multiple raters in dichotomous classification tasks, incorporating both BM and BT. Since BM and BT are unobservable in real-world settings, two sequential Bayesian models using Markov Chain Monte Carlo methods are introduced to estimate them. The first model infers each rater's BT based on observed classifications and a prior on BMs. The second model uses the estimated BTs to infer BMs. Estimated BMs are then dichotomized using a fixed $BT = 0.5$ to derive adjusted classifications, which are used to compute the adjusted reliability indices. Two simulation studies were conducted to test the performance of the adjusted indices compared to their traditional versions. Results show that the adjusted indices consistently outperform their standard counterparts, particularly when BT differences among raters are large. These improvements are observed across all simulated conditions and all three indices, highlighting the robustness and generalizability of the method. Importantly, the approach is fully applicable post hoc, requiring only standard binary classification data, and provides researchers with nuanced insights into inter-rater reliability.

Stability rate and a dynamic stopping criterion in KST-based adaptive assessment

Luca Stefanutti, Debora de Chiusole, Pasquale Anselmi
University of Padua

In the mid-1980s, the theory of knowledge structures was developed with a key goal: to provide a formal foundation for building an "efficient machine for the assessment of knowledge" (Doignon & Falmagne, 1999, Knowledge Spaces, Springer). This forward-looking vision has recently found practical applications in cognitive and neuropsychological assessment, raising new questions about the suitability of adaptive methods in such contexts. Let Q be the full set of test items. A key challenge is that in adaptive assessment, only a subset $A \subseteq Q$—the assessment set—is administered, and it may differ between individuals. This variability introduces a degree of non-standardization. A natural question follows: what is the probability that the outcome based on A matches that based on the full set Q? If this probability is 1, the shorter assessment is fully equivalent to the complete one. A widely used algorithm in this domain is the Continuous Markov Procedure (CMP), which infers the most likely knowledge state—the mode of a posterior likelihood—based on responses. Denote this state by $\hat{K}_A$ for A, and by $\hat{K}_Q$ for Q. The stability rate is the probability that $\hat{K}_A = \hat{K}_Q$. This talk presents the derivation of that rate. If, after answering A, this probability is 1, the full set Q is unnecessary. This insight leads to a novel stopping rule: terminate the CMP once the stability rate exceeds a predefined threshold (e.g., 0.90). Simulations and real-data analysis show that this criterion significantly enhances CMP efficiency, matching standard accuracy while requiring fewer items.

Combining theory- and data-driven methods to build Knowledge Spaces: The case of linear functions

Peter Steiner, Jan Hochweber, Michael Kickmeier-Rust,
Stephanie Leininger
University of St. Gallen

To optimally support student learning, teachers need informative assessments of students' current state of knowledge. However, many practitioners find that conventional assessment methods fall short in informing teaching meaningfully (Huff & Goodman, 2007). To address this issue, the project "Investigating the Spaces and Trajectories in Mathematics Learning" (SaTiM), funded by the Swiss National Science Foundation (SNSF), seeks to develop an assessment tool for the domain of linear functions using the statistical framework of Competence-based Knowledge Space Theory (CbKST; de Chiusole et al., 2020), which allows to conduct assessments within a Knowledge Space (KS). This approach allows to derive next feasible learning steps directly from the assessment results. In the initial conceptual phase, in which top-down (systematic literature review) and bottom-up (item analysis) approaches were combined iteratively, a domain consisting of 119 competencies was identified. In the current project phase, this domain is being structured into a KS by identifying surmise relations among these competencies. To this end, different approaches can be found in the literature, that can be categorized into data-driven and theory-driven approaches. Whereas data-driven approaches have received more attention in the last decade, they require extensive item response data. Consequently, theory-driven approaches, which mostly rely on expert queries, may be more feasible for projects without preexisting datasets. The QUERY algorithm and its extension PS-QUERY (Cosyn & Thiéry, 2000) are the most prominent frameworks to elicit these queries. Their main limitations are the number of necessary expert queries and the resulting careless errors caused by exhaustion. In the SaTiM project, preexisting datasets are available only to a limited extent. Given the limitations of the query approach, a combined strategy incorporating qualitative techniques and data-driven refinement is currently being conceptualized. In this contribution, existing data-driven and query-based approaches for building knowledge spaces will be summarized, and the combined methodology proposed for the application in SaTiM will be discussed. As there is, to our knowledge, a lack of available software to conduct expert queries within modern programming languages or platforms, an open-source web application is currently being developed and presented as part of our contribution.

Generalized Bayesian hierarchical structural equation modeling

Niek Stevenson, Andrew Heathcote, Dylan Molenaar, Birte Forstmann, Dora Matzke
University of Amsterdam (NL), University of Newcastle (AU), University of Leiden (NL)

Structural equation models (SEMs) are popular tools for investigating structural relationships among latent psychological constructs. In traditional applications, SEMs are estimated on summary scores aggregated across multiple measurements per individual, ignoring the hierarchical structure of the data and assuming that the individual-level data are normally distributed. This approach suffers from two shortcomings. First, aggregate scores do not account for measurement error in the individual-level data, leading to smaller estimates of the structural relationships due to unaccounted for variability. Second, the assumption of normality is implausible in many applications and fails to provide a substantive psychological account of the processes that give rise to the data. Here we propose a Bayesian hierarchical SEM framework that addresses both limitations. Our approach allows researchers to flexibly model the individual-level data, ranging from the traditional normal distribution to generative cognitive models, such as evidence-accumulation or reinforcement learning models. The joint hierarchical estimation of the individual-level model, and the structural relationships among the latent constructs extracted from it, takes into account measurement uncertainty and enables the use of Bayesian model selection techniques. We showcase how hierarchical SEMs can be used to test the latent structure of psychological constructs extracted from a psychological model fit to data obtained from multiple conditions, tasks, or data streams.

A hitchhiker's guide to information theoretical measures in psychology

Niels Van Santen, Yves Rosseel, Daniele Marinazzo
Ghent University

In psychology, as in other sciences, information theory can be used as a tool to complement more standard regression-based methods of data analysis. It is important to see the potential of information theoretical measures as statistical tools without implying a connection to their origins in communication theory and engineering. The use of these measures may provide us with additional insights due to their sensitivity to non-linear relationships, their flexibility to the mixing of data types, and their more straightforward generalization towards investigating higher-order interactions. We briefly reintroduce information theory and compare several measures such as mutual information and co-information with correlation and regression-based methods for the investigation of variable dependence.

A computational framework for integrating Predictive processes with evidence Accumulation Models (PAM)

Antonino Visalli[a], Francesco Maria Calistroni[b], Margherita Calderan[b], Francesco Donnarumma[c], Marco Zorzi[b], Ettore Ambrosini[b]
[a] *University of Modena and Reggio Emilia,* [b] *University of Padova,* [c] *National Research Council of Italy*

Evidence Accumulation Models (EAMs) have been widely used to investigate speeded decision-making processes, but they have largely neglected the role of predictive processes emphasized by theories of the predictive brain. We introduce the "Predictive evidence Accumulation Models" (PAM), a novel computational framework that integrates predictive processes into EAMs. Grounded in the "observing the observer" framework, PAM combines models of Bayesian perceptual inference, such as the Hierarchical Gaussian Filter, with three established EAMs (the Diffusion Decision Model, Lognormal Race Model, and Race Diffusion Model)

to model decision-making under uncertainty. We validate PAM through extensive parameter recovery simulations, demonstrating its accuracy, robustness, and computational efficiency across varied decision-making contexts. To facilitate adoption, we will present a concise, step-by-step tutorial using real behavioral data and discuss PAM's theoretical and practical implications for modeling uncertainty and predictive cognition. PAM advances the integration of predictive brain theories and decision modeling, offering a unified and extensible tool for empirical research. All code, data, and additional examples (including multiple perceptual models and programming languages) are openly available at osf.io/3jve9 and github.com/antovis86/PAM-PredictiveAccumulationModels.

List of Authors

September 3th, 2025

08:30	09:00	*Opening session*
09:00	**10:00**	***Keynote: Juergen Heller***
		KNOWLEDGE SPACE THEORY 1
10:00	10:20	*Jenisch A.*
10:20	10:40	*de Chiusole D.*
10:40	11:00	*Anselmi P. (chair)*
11:00	11:30	*Coffee Break*
		PERCEPTION AND PSYCHOPHYSICS
11:30	11:50	*Balakrishnan J.*
11:50	12:10	*Hsu Y.-F.*
12:10	12:30	*Kohler D. (chair)*
12:30	14:00	*Lunch*
		STATISTICAL METHODS
14:00	14:20	*Alfarone I.*
14:20	14:40	*Arletti A.*
14:40	15:00	*Guerrieri V.*
15:00	15:20	*Manente M.*
15:20	15:40	*Spoto A. (chair)*
15:40	16:10	*Coffee Break*
		PSYCHOMETRICS
16:10	16:30	*Flebus G.*
16:30	16:50	*Mignemi G.*
16:50	17:10	*Epifania O.M. (chair)*

September 4th, 2025

09:00	**10:00**	***Keynote: Dora Matzke***
		COGNITIVE MODELS
10:00	10:20	*Donzallaz M.*
10:20	10:40	*Petrin A.*
10:40	11:00	*Heathcote A. (chair)*
11:00	11:30	*Coffee Break*
		KNOWLEDGE SPACE THEORY 2
11:30	11:50	*Brancaccio A.*
11:50	12:10	*Gondan M.*
12:10	12:30	*Stefanutti L.*
12:30	12:50	*Noventa S. (chair)*
12:50	14:00	*Lunch*
		DECISION MAKING
14:00	14:20	*Diederich A.*
14:20	14:40	*D'Ambrogio S.*
14:40	15:00	*Seitz F.I.*
15:00	15:20	*Regenwetter M.*
15:20	15:40	*Doignon J.-P. (chair)*
15:40	17:00	*Coffee Break & Poster session*
17:00	18:00	*Business Meeting*
19:30		*Social Dinner*

September 5th, 2025

09:00	**10:00**	***Keynote: Michel Regenwetter***
		MEASUREMENT AND SCALING
10:00	10:20	*Angelelli M.*
10:20	10:40	*Iurev R.*
10:40	11:00	*Dzhafarov E. (chair)*
11:00	11:30	*Coffee Break*
		BAYESIAN METHODS
11:30	11:50	*Edmar B.A.*
11:50	12:10	*Rioul O.*
12:10	12:30	*Stevenson N. (chair)*
12:30	14:00	*Lunch*
		REACTION TIMES
14:00	14:20	*Legner L.*
14:20	14:40	*Lombardi L.*
14:40	15:00	*Hadian Rasanan A.H.*
15:00	15:20	*Visalli A.*
15:20	15:40	*Colonius H. (chair)*
15:40	16:00	*Closing session*