# IOPS & SMiP Summer Conference
# June 23-24, 2022 Leuven

## Abstract and Program booklet

# Program Overview

**Program Thursday 23 June**

**10.00 – 10.30   Receipt with coffee and tea**

10.30 – 10.45   Official Opening
Welcome words from IOPS (Rob Meijer),
SMiP (Thorsten Meiser) &
KU Leuven (Francis Tuerlinckx)

*Morning session - Chair: Rob Meijer*

10.45 – 11.15   *Theresa J.S. Koch (University of Mannheim)*
Psychological Detachment Matters Right After Work:
Engaging in Physical Exercise After Stressful Workdays

11.15 – 11.45   *Constantin Meyer-Grant (University of Freiburg)*
Disentangling different aspects of between–item similarity
unveils evidence against the ensemble model of lineup memory

11.45 – 12.15   *Sebastian Castro-Alvarez (University of Groningen)*
Developing an Item Response Theory Model to Analyze
Psychological Time Series: A rollercoaster of failures and
successes

**12.15 – 13.15   Lunch**

*Afternoon session - Chair: Thorsten Meiser*

13.15 – 13.45   *Evelien Schat (University of Leuven)*
Prospectively detecting mean and variance changes
through statistical process control

13.45 – 14.15   *Ester Maassen (University of Tilburg)*
Widespread Disregard of Measurement Invariance in Comparing
Psychological Scales across Groups, Conditions, and Time
Points

**14.15 – 14.30   Coffee Break**

14.30 – 15.00   *Marcel Schreiner (University of Landau)*
Agentic binding: Influences of agency on binding processes in
episodic memory

15.00 – 15.30   *Rebekka Kupffer (University of Landau)*
Careless responding in multidimensional forced–choice
questionnaires: what does it look like and how can it be
detected?

**15.30 – 16.00   Coffee Break**

16.00 – 17.00   *Invited presentation: Sylvia Wenmackers*
                *(Centre for Logic and Philosophy of Science, University of Leuven)*
                **Philosophical Introduction to**
                **inductive inference and confirmation theory**


17.00 – 18.30   **Poster Session & Reception**

**A tool to simulate and visualize dyadic interaction dynamics**
                                *Sophie Berkhout (University of Utrecht)*

**Momentary profile similarity measures for multivariate dyadic time series**
                                *Chiara Carlier (University of Leuven)*

**A Multi-Method-Approach for Identifying Ethical Concerns Influencing the Acceptance of Emerging Technologies in Real Time**
                                *Julius Fenn (University of Freiburg)*

**A signal recognition model of intensity processing**
                                *Robert Johansson (University of Tübingen)*

**It's all about timing: Exploring the consequences of choosing different temporal resolutions for analyzing passive measures**
                                *Anna Langener (University of Groningen)*

**Testing sampling-based models of probability judgments with an event ranking task**
                                *Xiaotong Liu (University of Mannheim)*

**Validating IRT model parameters by jointly modeling item responses and response times**
                                *Viola Merhof (University of Mannheim)*

**How best to quantify replication success? a simulation study on the comparison of replication success metrics**
                                *Jasmine Muradchanian (University of Groningen)*

**Just in time: A replication and storage-retrieval analysis of the temporal gradient of retroactive interference in episodic memory**
                                *Julian Quevedo Pütter (University of Mannheim)*

**No more information, please! Bayesian modeling insights about sequential advice seeking**
                                *Tobias R. Rebholz (University of Tübingen)*

Predictive accuracy analysis: A new sample size planning method in the context of VAR(1) models

*Jordan Revol (University of Leuven)*

Examining qualitative differences in emotion differentiation by applying latent Markov factor analysis

*Marcel Schmitt (University of Landau)*

The dark factor of personality as a common basis for aversive maladaptive traits

*David Scholz (University of Landau)*

Neural Bayesian Method for Estimating Complex Dynamic Models of Cognition

*Lukas Schumacher (University of Heidelberg)*

Asymmetrical Reaction Time (RT) Disparity Among Congruent, Neutral, and Incongruent Trials in Conflict Tasks

*Parker Smith (University of Tübingen)*

Bayesian Evaluation of Approximate Measurement Invariance

*Dandan Tang (University of Utrecht)*

**18.30** **Conference dinner**
Best Paper of 2021 Award

**Program Friday 24 June**

**09.00 – 09.30**     **Receipt with coffee and tea**

*Morning session - Chair: Janne Adolf*

09.30 – 10.00     *Damiano D'Urso (University of Tilburg)*
The effect of acquiescence bias on measurement invariance testing

10.00 – 10.30     *Adam Finnemann (University of Amsterdam)*
The urban desirability paradox? A large scale study of UK urban psychology

**10.30– 10.45**     **Coffee Break**

10.45 – 11.15     *Ruben Laukenmann (University of Mannheim)*
Did you see a weapon? – Is racial bias in weapon identification due to illusory perception or response execution failure?

11.15 – 11.45     *Sigert Ariens (University of Leuven)*
Including context and serial dependence in regression models of time series

**11.45 – 12.45**     **Lunch**

*Afternoon session - Chair: Irene Klugkist*

12.45 – 13.15     *Mihai Alexandru Constantin (University of Tilburg)*
A General Monte Carlo Method for Sample Size Analysis in the Context of Network Models

13.15 – 13.45     *Merijn Mestdagh (University of Leuven) – guest speaker*
A new era of dynamical measurement

**13.45 – 14.15**     **Coffee Break**          Board Meeting IOPS   (Room TI 01.05)

14:15 – 14:45     PhD Meeting          Joint Board Meeting IOPS – SMiP (Room TI 01.05)

14.45 – 15.15     Meeting IOPS & SMiP board and PhD representatives
Best Presentation and Poster Award
Closing Ceremony

# Philosophical introduction to inductive inference and confirmation theory

*Sylvia Wenmackers (University of Leuven)*

Philosophers of science have discussed different forms of scientific reasoning, including inductive inference. They have searched for a "logic of induction" and aimed to formalize when and how evidence confirms a theory or hypothesis - and by how much.

This talk gives an overview of the various proposals for a theory of confirmation and their challenges; this includes but is not limited to Bayesian confirmation theories. It also covers the current view by philosophers on falsifiability, corroboration, and risky predictions: concepts that were central to the work of Karl Popper and that continue to be influential in thinking about scientific methods.

# Psychological Detachment Matters Right After Work: Engaging in Physical Exercise After Stressful Workdays

*Theresa J.S. Koch (University of Mannheim)*

Whilst regular physical exercise as a recovery activity has an important role regarding employees' well-being, employees seem to engage less in physical exercise after stressful workdays. This paradoxical pattern between job stressors, recovery, and well-being has also been described as the recovery paradox. Aiming for a better understanding of this recovery paradox applied to physical exercise, we investigate psychological detachment right after work as an explaining psychological mechanism. We collected data from 97 employees on a total of 537 days using a daily diary study over two consecutive workweeks. Results of our two-level path model suggest that psychological detachment right after work indeed serves as an explaining mechanism in the paradoxical pattern between job stressors (workload, self-control demands) and time spent on physical exercise. However, we did not find support for a serial indirect effect from job stressors via psychological detachment and time spent on physical exercise to next-morning well-being (vigour, fatigue). Our study suggests a new perspective on the temporal sequence of psychological detachment and physical exercise by demonstrating psychological detachment as a prerequisite instead of an outcome of physical exercise after stressful workdays. This new perspective might be considered in future research and practical interventions.

Student discussant: Damiano D'Urso

# Disentangling different aspects of between-item similarity unveils evidence against the ensemble model of lineup memory

*Constantin Meyer-Grant (University of Freiburg)*

In order to model recognition decisions in the context of simultaneous eyewitness-identification lineups, two models based on signal detection theory—the independent observations and the ensemble model—are currently under consideration. These two models mainly differ with respect to their assumptions regarding the interplay between the memory signals of different stimuli presented in the same lineup. The independent observation model, on the one hand, assumes that the decision maker separately assesses the memory signal of each simultaneously presented test stimulus. The ensemble model, on the other hand, assumes that each of these memory signals is first compared with and then assessed relative to its respective context (i.e., the memory signals of the other stimuli within the same lineup). Here we theoretically demonstrate that the independent observations model can actually also produce a data pattern that is frequently cited as evidence for the ensemble model (viz. the dud-alternative effect). We then introduce a cleaner test of these models by including a lure (i.e., a non-studied item) in a lineup that resembles a studied item other than the target of the current lineup. This illustrates that the similarity between a new and an old item (old–new similarity) and the similarity of items belonging to the same lineup (within-lineup similarity) should be considered as two distinct concepts and should be modeled accordingly, which has not yet been done in eyewitness-identification research.

Student discussant: Sigert Ariens

# Developing an Item Response Theory Model to Analyze Psychological Time Series: A rollercoaster of failures and successes

*Sebastian Castro-Alvarez (University of Groningen)*

Nowadays, many researchers are collecting intensive longitudinal data to study psychological dynamics. To analyze this kind of data, researchers have relied on diverse methods ideal for time series. However, most of these methods assume that the variables in the time series are measured on an interval scale, which is not the case when Likert-scale items are used. This can be problematic because ignoring the scale of the variables can bias the results. To overcome this, we developed an IRT model for N=1 time series data, which is called the time-varying dynamic partial credit model (TV-DPCM). The model integrates the partial credit model and the time varying vector autoregressive model. Furthermore, as the model was implemented within the Bayesian framework, we also proposed posterior predictive model checking methods to assess the fit of the TV-DPCM. We illustrate how to use and interpret the model with empirical data from one subject. In general, the proposed model seems to be a promising method to further understand psychological measurement in intensive longitudinal data.

Student discussant:          Sophie Berkhout

# Prospectively detecting mean and variance changes through statistical process control

*Evelien Schat (University of Leuven)*

Retrospective analyses of experience sampling (ESM) data have shown that changes in mean and variance levels may serve as early warning signals of an imminent depression. Detecting such early warning signs prospectively would pave the way for timely intervention and prevention. The exponentially weighted moving average (EWMA) procedure seems a promising method to scan ESM data for the presence of mean changes in real-time. Affective ESM data, however, violate assumptions of the EWMA procedure: the observations are not independent across time, often skewed distributed and characterized by missingness. To deal with these data characteristics, we compute and monitor the day averages rather than the individual measurement occasions with EWMA. Since simulation studies showed promising results, we use a similar approach for the detection of variance changes, where we compute and monitor day statistics of variability (i.e., variances, standard deviations and the natural logarithm of the standard deviations). Simulation studies again show good performance, and allow to provide recommendations on which statistic of variability to monitor based on the type of change (i.e., variance increase or decrease) one expects. This is good news because existing exponentially weighted procedures for monitoring variability directly (e.g., EWMV, EWMA-S²) all have important downsides, preventing easy application. Additionally, we reflect on the design choices of ESM studies, as they also influence the performance of the EWMA procedure. Specifically, the number of beeps per day (i.e., sampling frequency), the distribution of the in-control data and the number of in-control days influence the performance of EWMA.

Student discussant:　　　　Maximilian Linde & Leonie Cloos

# Widespread Disregard of Measurement Invariance in Comparing Psychological Scales across Groups, Conditions, and Time Points

*Ester Maassen (University of Tilburg)*

In psychological science, self-report scales are widely used to compare means in targeted latent constructs across time points, groups, or experimental conditions. For these scale mean comparisons to be meaningful and unbiased, the scales should be measurement invariant with respect to the compared time points or (experimental) groups. Measurement invariance testing allows a psychometric check if the latent constructs are measured equivalently across groups or time points. We sampled 426 psychology articles with open data that involved a total of 919 scale mean comparisons to (1) investigate common practices in conducting and reporting of measurement invariance testing, (2) computationally reproduce reported measurement invariance test results, and (3) conduct measurement invariance tests for 162 scale mean comparisons that enabled sufficiently powerful measurement invariance testing with the shared data. We find that measurement invariance tests with scale mean comparisons are uncommonly done and poorly reported in psychology, and most failures of measurement invariance characterize scale mean comparisons concerning experimental conditions. We offer recommendations on reporting standards for measurement invariance and on improving computational reproducibility practices.

Student discussant:  Dandan Tang & Nils Petras

# Agentic binding: Influences of agency on binding processes in episodic memory

*Marcel Schreiner (University of Landau)*

Memory representations in episodic memory are typically coherent, which requires the events' constituent elements to be bound together. Only a limited number of moderators of such binding processes have been identified. In five experiments we tested whether agency facilitates binding. The results strongly hinted at a facilitating effect of agency and additionally suggested an effect of presentation format, with binding effects only being found in case of simultaneous, but not sequential, event element presentation. This suggests that the presence of an agentic element in an event facilitates the formation of coherent memory representations, which may be a more proximate explanation of previously found effects of animacy on the binding of event elements. The results further add to previous literature suggesting that additional processes may be required when binding event elements across several temporarily divided encoding episodes than when binding them within a single encoding episode.

Student discussant:      An-Chiao Liu

# Careless responding in multidimensional forced-choice questionnaires: what does it look like and how can it be detected?

*Rebekka Kupffer (University of Landau)*

Careless responding (CR) is a response behavior characterized by selecting response options without considering the (whole) item content. Currently, little is known about how CR manifests itself in questionnaires in the multidimensional forced-choice (MFC) format. In a laboratory study (N = 430), we manipulated CR and compared its manifestation on eleven CR measures. Participants were assigned to one of three conditions: a control condition ($n_1$ = 140) and two CR conditions in which we either instructed participants to respond carelessly ($n_2$ = 140) or distracted them to induce CR ($n_3$ = 150). The experiment consisted of a survey with five personality inventories in the MFC format and an oral interview. To detect CR, we included self-report items and instructed response triplets in the survey. Moreover, we calculated post-hoc indices including analyses on response times and rank-order patterns as well as consistency and outlier analyses. We found significant differences between the control condition and the instructed CR condition on all of the indices, except for response times. However, the control and the distracted responding conditions did not differ significantly. In an exploratory comparison of the instructed and distracted conditions, we found small to large differences on most indices. Commonly mentioned strategies when being listless, tired or unmotivated were to read the items superficially, think less about the item content, choosing a random rank order, or copying the presented one. Follow-up studies will investigate the validity of the CR indices and adequate cutoff values. An outlook on these studies will be given.

Student discussant:        Sebastian Castro-Alvarez & Jasmine Muradchanian

# The effect of acquiescence bias on measurement invariance testing

*Damiano D'Urso (University of Tilburg)*

In social sciences, group differences concerning latent constructs (e.g., self-esteem) are ubiquitously investigated, and these constructs are often measured using scales composed of multiple self-report ordinal items. For these comparisons to be valid, it is fundamental that these measures function equivalently across groups, or, in technical jargon, measurement invariance (MI) must hold. Testing for MI allows one to investigate group-specific systematic biases that occur in item responses across groups. A potential source of systematic bias in self-report measures is that of response styles or response bias, which can be viewed as a stylistic tendency in the manner respondents use a rating scale when responding to self-report items. Acquiescence response style (ARS) is a well-known response bias, which represents a tendency to agree with items regardless of their content. Failing to take into account an ARS when testing for MI may result in concluding that a measure is non-invariant while this is purely due to this stylistic tendency. In this project, by means of a simulation study, we investigated the effect of ARS on MI testing both when such tendency is taken into account by including an additional ARS factor in the measurement model (MM) or not. Based on the simulation study results, recommendations and guidelines are provided for applied researchers.

Student discussant: Jasmine Muradchanian & Jeffrey Durieux

# The urban desirability paradox? A large scale study of UK urban psychology

*Adam Finnemann (University of Amsterdam)*

With the majority of the global population living in urban environments it is paramount to ensure that city life is healthy and sustainable. To this end, we present a broad study of urban psychology using the UK Biobank data of 157k participants.

The study is inspired by recent literature suggesting that, in developed countries, the psychological well-being of the urban population is lower than the rural one . This is named the rural happiness paradox due to its contrast with cities' popularity as well as social and economic opportunities. We aim to explore this phenomenon further and some drivers of it.

To do so we present a cross-sectional study of the relation between urbanicity, social satisfaction, economic satisfaction, and psychological well-being. We use polynomial regression to explore the non-linear associations between urbanicity and aforementioned variables in a bottom-up manner. Employing this definition, we assess how (1) mean-levels, (2) variability (inequality level), and (3) correlations between happiness and socio-economic variables (importance level) change as a function of urbanicity.

Our preliminary results indicate the UK cities pose a threat to psychological well-being. Despite the wealth, opportunities, and popularity city residents score worse on all psychological dimensions. We name this pervasive concern the urban desirability paradox. Our results indicate that increased inequality and financial needs are contributing drivers of the effect.

Student discussant: Xinkai Du & Sophie Berkhout

# Did you see a weapon? – Is racial bias in weapon identification due to illusory perception or response execution failure?

*Ruben Laukenmann (University of Mannheim)*

The Weapon Identification Task (WIT) is a sequential priming paradigm, which assesses the effect of racial priming (i.e., White vs. Black males) on visual discrimination between weapons (e.g., a gun) and innocuous objects (e.g., a tool). A typical finding is that participants tend to make more errors in identifying innocuous objects after seeing Black versus White male face primes. Regarding the cause of this identification error two competing mechanisms have been proposed: the illusory perception hypothesis and the execution failure hypothesis. The illusory perception hypothesis states that perception of the targets is colored early on by the preceding prime's race. For example, a metal tube as part of an object might more easily be interpreted as part of a gun if preceded by a Black male face prime and more easily as part of a tool (e.g., a screwdriver) if preceded by a White male face prime. In contrast, the execution failure hypothesis states that threat-stereotype based associations elicited by primes interfere in participants response execution. Two studies applying a base rate manipulation for target objects in the WIT indicate support for the execution failure hypothesis. This even holds true for explicitly instructing participants to use race in target identification. These results suggest that participants perceive the target objects correctly but fail to execute the correct response due to interfering threat-stereotype associations. Hence strategies increasing participants abilities to impede the influence of automatic associations on response execution may help to improve participants' performance in the WIT and reduce biased responding.

Student discussant: Jordan Revol

# Including context and serial dependence in regression models of time series

*Sigert Ariens (University of Leuven)*

To model the dynamics present in time series of psychological data, researchers often appeal to autoregressive structures. In recent years, the need to take into account contextual influences, e.g. emotionally relevant events in the study of affect dynamics, has become recognized. A straightforward way of doing so is by including covariates in the model equation. A first option is to include covariates in autoregressive models. Alternatively, one can include covariates in classical linear regression models, allowing for autocorrelation in the residuals. Researchers in the behavioral sciences have proposed situations in which one of the two approaches should be preferred, e.g. depending on whether contemporaneous or lagged effects are of prime interest. However, evidence for these propositions remains inconclusive. In this talk, we delineate the differences between the two approaches, and hope to provide the information needed for researchers to make an informed choice on this decision. Specifically, we show that the residualized approach can impose implicit restrictions on the model-implied relationships between the variables, which can result in biased estimates of the model parameters if these restrictions do not hold for the data at hand. We also touch on how the restrictions can be tested, relying on a simple likelihood-ratio test procedure, also for multilevel versions of the considered models. We present the results of a simulation study confirming these insights, and provide a real-data example showing that misleading results can be obtained when the residualized approach is used without testing the restrictions invoked by the model structure.

Student discussant:     Julius Fenn & Daan de Jong

# A General Monte Carlo Method for Sample Size Analysis in the Context of Network Models

*Mihai Alexandru Constantin (University of Tilburg)*

The network approach to psychology is an increasingly popular framework for studying pairwise interactions among variables. As the field matures and psychological network modeling becomes more prevalent, there is an increasing need to aid researchers with a network approach in mind that plan to collect data. In this talk, I introduce a general method for performing sample size analysis in the context of network models. The method takes the form of a three-step recursive algorithm designed to find an optimal sample size value given a model specification, an outcome measure (e.g., sensitivity), and a statistic of interest (e.g., power). It starts with a Monte Carlo simulation step for computing the outcome measure and the statistic at various sample sizes. It continues with a monotone non-decreasing curve-fitting step for interpolating the statistic. The final step employs stratified bootstrapping to account for the uncertainty around the interpolated curve. In the first part of this talk, I provide an overview of the method and discuss its validation and performance. In the second part, I illustrate, in the form of a tutorial, how the method can be applied to a network model. The tutorial showcases the open-source implementation of the method as an R package called powerly.

Student discussant:       Letty Koopman

# A new era of dynamical measurement

*Merijn Mestdagh (University of Leuven)*

In order to reach robust conclusions about people's real-life emotional or symptom fluctuations, and to truly advance our understanding of the temporal regularities that underlie our experiences, a first prerequisite is the reliable assessment of people's momentary emotions and symptoms. Unfortunately, data gathered with the experience sampling method is often clouded by measurement noise. In this presentation we discuss how we can boost the signal to noise ratio of an experience sampling research protocol by making the protocol more dynamic. Specifically we introduce three tweaks, currently possible in the m-Path experience sampling platform. First we discuss how experience sampling bursts can be programmed to measure the moments that matter in a higher frequency. Second, we discuss how to make measurements relative instead of absolute. Finally, we discuss how to change the content of questions conditional on (a combination of) previous answers.

# A tool to simulate and visualize dyadic interaction dynamics

*Sophie Berkhout (University of Utrecht)*

Dynamic models are becoming increasingly popular to study the dynamic processes of dyadic interactions. I present simulations and visualizations of data to illustrate the model-implied processes of various statistical models that can be used to analyze dyadic data, and elaborate on specific features of these models. Specifically, I consider: (a) the first-order vector autoregressive (VAR(1)) model; (b) the latent VAR(1) model; (c) the time-varying VAR(1) model; (d) the threshold VAR(1) model; (e) the hidden Markov model; and (f) the Markov-switching VAR(1) model. Furthermore, to make data generation for these various models accessible to a wide audience, I provide a Shiny web application that allows researchers to easily generate data from the models discussed through a graphical user interface. The app is available at: https://utrecht-university.shinyapps.io/dyadic-interaction-dynamics/. With this tool, researchers can alter parameter values, inspect what kind of patterns this generates using various data visualizations, and also download the generated data to use for further analyses. Our goal is to make these statistical models more accessible to relationship researchers so that they can make more informed decisions on what modeling approach fits their research question and data best.

# Momentary profile similarity measures for multivariate dyadic time series

*Chiara Carlier (University of Leuven)*

People live in continuous relation to each other and their environments, forming dyads: college roommates, romantic partners, parents, colleagues, ... These social relationships are of major importance in the course of our lives and impact our well-being both physically and mentally. One important or even essential quality to help us understand social relationships is how similar two people are feeling or behaving. Until now, similarity has mostly been examined from a variable-centered approach using bivariate measures and from a cross-sectional approach, using single measurements or aggregate values. However, human interactions take on unique forms and fluctuate with varying contexts. Therefore, a multivariate, dyad-centered, longitudinal approach is better suited for dyadic interactions. In consequence, this requires methods to capture similarity in multivariate dyadic time series. Departing from the existing profile similarity methods, we developed momentary profile similarity measures. These measures take the profile similarity between the state ratings of two people at a specific moment and can then be looked at over a given time course. We will guide the audience through several steps of this methodology by applying the momentary measures to an existing longitudinal data set of discrete emotions, rated by romantic couples. We will show how to compute two different momentary profile similarity measures, explore their distribution, and relate these measures to couple and partner covariates.

# A Multi-Method-Approach for Identifying Ethical Concerns Influencing the Acceptance of Emerging Technologies in Real Time

*Julius Fenn (University of Freiburg)*

Emerging technologies are at an early research and development stage and no final products are available yet. As such, these technologies are not integrated in existing normative frameworks for decision-making ("business-as-usual situation"). Without established legal and ethical guidelines, emerging technologies can cause massive uncertainties: Often decisions, like the deployment of technologies in specific contexts, are required while information is missing and the possible effects as well as side-effects and potential risks of deploying technologies are hardly predictable. To reduce uncertainty, it is necessary to make informed but incremental decisions in "real time" and to ethically assess emerging technologies throughout the entire development process.

To facilitate the ethical assessment of emerging technologies, we have developed an "Ethics Scale of Technology Assessment", derived from multiple moral theories based on an extensive scoping review. The scale consists of 48 items belonging to six moral theories, yet a preliminary study favored a one-factor solution. In addition, to assess possible additional influential factors, Cognitive-Affective Maps (CAMs; a tool to visualize attitudes and beliefs), can be collected using our browser-based software (C.A.M.E.L.) and analyzed with its dedicated software (CAM-App). As major result, this multi-method-approach allows to (1) discover a broad range of ethical concerns influencing the acceptance of emerging technologies, and (2) potentially identify further relevant aspects that are not considered in our developed scale.

# A signal recognition model of intensity processing

*Robert Johansson (University of Tübingen)*

A model describing how people categorize sensory events of different stimulus intensity is advanced. The central idea draws inspiration from mathematical conceptions of the sensory encoding stage according to which the output of peripheral stimulus transducers is a stream of afferent neural pulses travelling towards a central decisional loci.

The bulk temporal properties of the neural time series is assumed to be broadly captured by the stationary Poisson process. In the model, the Poisson conjecture of sensory encoding is bridged with a sequential decision algorithm based on iterative likelihood ratio computations for the entire series of inter-pulse times recorded at decision center. Overarching features of the theory are brought to bear upon empirical data gathered in a brightness classification task.

It is demonstrated that a novel prediction of the present theory is a generalization of Pieron's law from simple reactions to the domain of binary choices. The results support a role of adaptive perception in intensity processing, according to which the relation between stimulus intensity and sensation magnitude ultimately depends on contextual factors such as the range of intensity levels contained within a stimulus set. Due to kinship between the advanced theoretical framework and Signal Detection Theory, the present work holds promise of unifying classical principles of signal detection with chronometrically oriented approaches to sensory-cognitive processing.

# It's all about timing: Exploring the consequences of choosing different temporal resolutions for analyzing passive measures

*Anna Langener (University of Groningen)*

Collecting data passively (e.g., through smartphones) offers great value to track (social) behavior by measuring individuals frequently over time with a low participant burden. Even though those passive measures are increasingly used, there are important methodological challenges that may hamper widespread uptake. A crucial one is the issue of time scale: the temporal resolution chosen to summarize the variables may have an impact on how results are interpreted. Despite its importance, the choice of temporal resolution is rarely justified. We address the various time-related decisions researchers face when analyzing passive measures and illustrate how choosing different time scales would lead to different substantive conclusions.

# Testing sampling-based models of probability judgments with an event ranking task

*Xiaotong Liu (University of Mannheim)*

People's explicit probability judgments often appear to be probabilistically incoherent. The most prominent example of this is the conjunction fallacy (Kahneman & Tversky, 1983). Recently, various sampling-based models have been proposed to account for different biases and fallacies in probability judgments. To date, these models have only been investigated in the probability estimation tasks (e.g., Costello & Watts, 2014). We propose an event ranking task to investigate sampling-based models of probability judgments. On each trial of the ranking task, participants will be asked to provide a ranking for an event set consisting of four events, A, not-A, B, and not-B, in terms of their perceived likelihoods. We formally derived qualitative predictions for this task by assuming direct sampling from the fixed underlying probability distributions. We also showed that adding read-out noise in the sampling process, as suggested in the Probability Theory plus Noise model (Costello & Watts, 2014), will not vastly change the qualitative predictions. In two online experiments, we asked participants to rank 12 different event sets and found evidence entirely in line with the qualitative predictions we derived.

# Validating IRT model parameters by jointly modeling item responses and response times

*Viola Merhof (University of Mannheim)*

Whenever respondents are asked to provide subjective information about themselves by responding to a Likert-type item, they are faced with the challenge of selecting one of the provided categories. Sometimes such decisions may come about quickly, sometimes respondents may be uncertain and take more time for the response selection. Here it is shown how the simultaneous modeling of item responses and response times can be used to validate IRT model parameters and how a joint model can provide new insights into cognitive processes involved in item responding. To this end, responses to items of a personality inventory were analyzed using a multi-process IRT model accounting for response styles. The log-transformed response times were analyzed by linear mixed modeling. Predictor variables included response style-related properties of the given responses, of the respondents, and their interactions. Further, person-item interactions were considered. Response times were consistent with the hypothesized pattern and supported the substantive interpretation of IRT model parameters.

# How best to quantify replication success? a simulation study on the comparison of replication success metrics

*Jasmine Muradchanian (University of Groningen)*

To overcome the frequently debated crisis of confidence, replicating studies is becoming increasingly more common. Multiple frequentist and Bayesian measures have been proposed to evaluate whether a replication is successful, but little is known about which method best captures replication success. This study is one of the first attempts to compare a number of quantitative measures of replication success with respect to their ability to draw the correct inference when the underlying truth is known, while taking publication bias into account. Our results show that Bayesian metrics seem to slightly outperform frequentist metrics across the board. Generally, meta-analytic approaches seem to slightly outperform metrics that evaluate single studies, except in the scenario of extreme publication bias, where this pattern reverses.

# Just in time: A replication and storage-retrieval analysis of the temporal gradient of retroactive interference in episodic memory

*Julian Quevedo Pütter (University of Mannheim)*

Retroactive interference in episodic memory has been shown to follow a temporal gradient: The longer the time interval between an original and an interpolated study phase, the better the performance in a final recall test. This temporal gradient of retroactive interference (TGRI) can be easily explained by temporal distinctiveness theory: According to this theory, the retrievability of an item from memory is a function of its temporal discriminability. In contrast, in the neuroscientific literature, the TGRI is considered as evidence for post-encoding memory consolidation. To investigate the robustness of the TGRI, we conducted an online replication study. Here, we used the significance of the critical interaction between study (original vs. replication study) and condition (short vs. long time interval between original and interpolated study phase) as an innovative criterion for replication success. In a second study, we will use a multinomial processing tree (MPT) model to test the predictions of the temporal distinctiveness and the consolidation account. More specifically, we will adjust the paradigm of our replication study to accommodate an adapted version of the Riefer and Batchelder (1995) storage-retrieval model. This model will enable us to disentangle the contributions of storage and retrieval processes to the TGRI. Thus, the parameter estimates from this model can be expected to allow for conclusions that have important theoretical implications for the validity of temporal distinctiveness theory and current consolidation theories.

# No more information, please! Bayesian modeling insights about sequential advice seeking

*Tobias R. Rebholz (University of Tübingen)*

In the dominant advice taking paradigm, the judge-advisor system, participants are provided with one single piece of advice. In more recent research, however, participants are allowed to freely sample multiple pieces of advice per judgment. At some point, the trade-off between the costs and benefits of sampling additional information tips in favor of stopping. For some participants, this tipping point is reached earlier than for others. The close correspondence between predicted and actual final judgments allows to derive conclusions with respect to participants' judgment formation and sampling behavior from a Bayesian updating perspective. For instance, as confidence grows at a decreasing rate along the sampling chain, the diminishing informational value of additional data renders stopping more and more likely. This observation also provides an explanation for belief perseverance by refusing to acquire further information due to the current level of confidence in a certain judgment.

# Predictive accuracy analysis: A new sample size planning method in the context of VAR(1) models

*Jordan Revol (University of Leuven)*

Researchers often use intensive longitudinal designs in combination with VAR(1) models to capture processes that evolve dynamically in time. In this context, sample size planning (i.e., number of measurement occasions needed) is in its early development and often uses power as criterion. One drawback is that power-based sample size recommendations will depend on the inspected model parameter at hand and will not hold for the model as a whole. Moreover, power analysis takes an explanatory stance while the predictive stance, that focuses on the performance of the full model for predicting unseen data, is increasingly used as well. We therefore suggest to consider predictive accuracy as a sample size planning criterion. Focusing on VAR(1) in a N=1 context, we propose a novel simulation-based method, called predictive accuracy analysis (PAA), to assess how many measurement occasions are required in order to optimize predictive accuracy. Specifically, we introduce a new predictive accuracy metric which is based on the VAR(1) assumptions, on the expected 'true' parameter values and on an acceptable size of the prediction errors, after standardizing them by computing the Mahalanobis distance between them and the true error distribution. This distance measure allows to account for the innovation covariances of the processes. Hence, our predictive accuracy analysis computes the sample size required so that the proportion of simulated replicates (e.g., .8) in which the inspected proportion of multivariate prediction errors of acceptable size is high enough. We then showcase how the different VAR(1) model parameters impact sample size planning recommendations.

# Examining qualitative differences in emotion differentiation by applying latent Markov factor analysis

*Marcel Schmitt (University of Landau)*

Emotion differentiation (ED) describes an individual's ability to describe their emotional experiences with specificity. To investigate ED, researchers typically use data from experience sampling studies in which participants repeatedly rate the intensity of multiple same-valenced emotions. ED has originally been conceptualized as a stable trait with the intra-class correlation coefficient across emotion ratings indicating an individual's general ED level. However, ED has increasingly been quantified on the momentary level and momentary ED indices have been derived from the overall trait ICC. One drawback of these quantitative ED indices is that they do not indicate which emotions in particular are differentiated from each other (and which are not) at a given moment in time. In order to enable a more qualitative insight into ED, we propose latent Markov factor analysis (LMFA) as an alternative modeling approach to investigate momentary ED. LMFA, a combination of latent Markov modeling and exploratory factor analysis, tracks changes in measurement models across time by clustering measurement occasions into latent states depending on the underlying measurement model. The factor loading pattern of the emotion items in a certain latent state could provide information on specific differentiation patterns across same-valenced emotions. Furthermore, the model can be extended to mixture LMFA by clustering individuals into latent classes according to the transition probabilities between latent states and time-constant and time-varying covariates can be added. In this poster, I present results from data of an experience sampling study in which we initially applied this modeling approach for the purpose of ED.

# The dark factor of personality as a common basis for aversive maladaptive traits

*David Scholz (University of Landau)*

In the alternative model for personality disorders of the DSM-5, so-called maladaptive traits have been proposed to describe dispositions that increase individuals' risk of psychosocial problems. As certain maladaptive traits do systematically covary, higher-order factors were inductively derived to describe these co-occurrences. However, at least in the case of the higher-order factor Antagonism, a dimension intended to describe clinically relevant tendencies which can be considered as socially or ethically questionable, an explanation of why the traits assigned to this dimension do co-occur is missing. Indeed, the definition of Antagonism is too vague for the classification of specific traits. Hence, to gain a better understanding of why maladaptive traits co-occur, we propose that the Dark Factor of Personality (D), which is by definition the core of all socially and/or ethically aversive tendencies, constitutes the underlying disposition from which certain maladaptive traits arise as manifestations. Thus, we test via structural equation modeling whether maladaptive traits, clearly describing socially and/or ethically aversive tendencies, can be seen as a manifestation of D, i.e. whether their aversive aspects are due to D. Additionality, we examine whether certain aspects of D, in turn, are insufficiently captured by aversive maladaptive traits and hence, are lacking in the DSM-5. Primarily data (N = 1781), using a bi-factor modeling approach, provides support to both propositions.

# A Neural Bayesian Method for Estimating Complex Dynamic Models of Cognition

*Lukas Schumacher (University of Heidelberg)*

Mathematical models of cognition are often memoryless and ignore potential fluctuations of their parameters. However, human cognition is inherently dynamic in nature, regardless of the reference time scale. Thus, we propose to augment mechanistic cognitive models with a temporal dimension and estimate the resulting dynamics within a superstatistics framework. To overcome the estimation challenges resulting from the complexity of superstatistical models, we develop and validate a simulation-based deep learning method for Bayesian updating which can recover both time-varying and time-invariant parameters. We then apply our method to estimate a dynamic implementation of the popular diffusion model from long time series of human response times data. We compare our results to those obtained by applying a benchmark diffusion model and show how the assumption of static or homogeneous parameters can preclude important temporal information.

# Asymmetrical Reaction Time (RT) Disparity Among Congruent, Neutral, and Incongruent Trials in Conflict Tasks

*Parker Smith (University of Tübingen)*

While the relation between congruent and incongruent conditions in conflict tasks has been the primary focus of cognitive control studies, the neutral condition is oft set as a baseline between the two conditions. Recent investigations have suggested that average neutral reaction times (RT) are not placed evenly between the two opposing conditions. The present study sought to establish whether there is a systematic asymmetry in RT differences between conditions in conflict tasks (Stroop, Flanker, and Simon Tasks). To explore this, we first conducted an exploratory meta-analysis that recorded the average RT's of conflict tasks with neutral conditions. Upon finding an asymmetry that showed a larger disparity between average neutral and incongruent RT's, we tested the prior mentioned conflict tasks with two different sets of stimuli in order to establish the robustness of the effect. The results suggest that the noted asymmetry does exist and may be dependent on stimuli and/or response mapping. More so, the asymmetry primarily appears in a manner that is opposite of what models such as the Diffusion Model of Conflict would suggest, leading to a possible extension of the model as a future direction to account for this.

# Bayesian Evaluation of Approximate Measurement Invariance

*Dandan Tang (University of Utrecht)*

Measurement invariance (MI) is of vital importance in multiple-group research with latent factors. In practice, exact MI is often not met, and approximate MI offers a working solution. This article thus proposes a Bayesian approach to test approximate MI in two-group confirmatory factor analysis (CFA) models using Bayes factors. This approach applies a new constraint on latent factors to make corresponding loadings or intercepts comparable across groups. These loadings or intercepts in two groups are about equal in the hypotheses of approximate MI. To test approximate MI, researchers have to carefully specify prior distributions for loadings and intercepts, as well as standardized tolerant difference that is the small difference and the corresponding loadings or intercepts are about equal across groups with this difference. Simulation studies explore prior choices and standardized tolerance differences. A flowchart of testing approximate MI is presented. The method is implemented in the function BMI() in the bain R-package, and a real data example was used to illustrate the procedure.