ABSTRACTS

PARALLEL PAPER SESSIONS

(Order: By number)

Dominik Ürüm

1.1 - Large Language Models as Translators of Thought: Automatic Coding of Verbal Reports for a Rule-Based Categorization Experiment

Dominik Ürüm , Technical University of Darmstadt

Frank Jäkel, Technical University of Darmstadt

Abstract:
proficiency in processing human language. This capability holds significant promise towards the automation of research in Cognitive Science, such as translating natural language into a programming language (Wong et al., 2023). In this work, we used LLMs for coding verbal reports from a two-choice rule-based categorization experiment (based on Shepard et al., 1961). While LLMs often produce correct translations and seemingly plausible outputs, they struggle with consistency and accuracy in certain edge cases — particularly those involving linguistically nested logic. This might introduce LLM-induced biases to the data, potentially obscuring insights into human behavior during subsequent analyses. We conclude that LLMs should not be used as a full replacement for human coders, but as a powerful component in a human-supervised pre-processing pipeline. This pipeline should include testing, such as format validation, consistency checks, and human reviewing to ensure the reliability of the research process. We demonstrate that this time-saving and systematic approach can yield new insights for rule-based categorization.

Steven Mcgannon & Anna-Mari Wallenberg

1.2 - LLMs, Instrumental Knowledge and Epistemic Agency

Steven Mcgannon , University of Helsinki

Tommi Buder-Gröndahl, University of Helsinki

Jesse Kuokkanen, University of Helsinki

Anna-Mari Wallenberg, University of Helsinki

Abstract:
How and to what extent do contemporary Large Language Models (LLMs) have knowledge? A recent discussion between Yildirim and Paul (2024a, 2024b) and Goddu et al. (2024) in Trends in Cognitive Sciences illustrates how the current philosophical disputes concerning the epistemic capabilities of LLMs often hinge on the conceptualization of agency and knowledge involved, and not on LLMs as such.

Yildirim and Paul claim that LLMs might have what they call instrumental knowledge, "...acquired through the successful use of instruments that perform certain tasks," (p. 405). For example, successful next word generation in LLMs would constitute such knowledge. For Goddu et al. this does not suffice for knowledge. According to them, only epistemic agents can have knowledge. As LLMs constitute mere models, they are not epistemic agents. Hence, they do not have knowledge.

Curiously, although the debate is framed in terms of LLMs, it isn't really about the capabilities or competencies of LLMs. In this paper, we argue that the debate hinges on several further assumptions about (i) the possibility of agency and knowledge in LLMs, (ii) the account of knowledge proffered, and (iii) the variety of epistemic tasks that require bona fide cognitive skills. Ultimately, we conclude that what matters in the debate is not about LLMs as such, but rather, how we conceptualize agency and knowledge.

In the discussion, Goddu et al. defend the view that LLMs do not possess instrumental knowledge. They only constitute models which actual epistemic agents (such as humans) can use for epistemic purposes. Following canonical philosophical epistemology, Goddu et al. hold that only epistemic systems can exhibit knowledge. For example, instruments such as measuring devices (thermometers etc.), are often taken to be akin testimony as indirect means to obtain knowledge (Sosa, 2006). Within a reliabilist virtue-epistemological framework, beliefs based on such information can attain the status of knowledge based on the reliability of these instruments. However, according to the canon, such considerations already require the system under discussion to be an epistemic agent, and thus, one might conclude, instrumental knowledge requires epistemic agency.

However, Goddu et al.'s position presupposes another, possibly more controversial, enactivist conception of epistemic agency. This conception of agency, familiar from the enactivist tradition of cognitive science (Varela et al., 1993; Hutto and Myin, 2013), maintains that only organic, biological and self-sustaining systems constitute agents – i.e. those that are autopoietic in the sense of Maturana and Varela (1980). In enactivist reasoning, epistemic

agency is only possible for such organic agents. As long as LLMs fall short of being agents in this sense, they cannot obtain knowledge.

Crucially, this enactivist presupposition entails that LLMs cannot obtain any type of knowledge no matter how well they perform in any range of linguistic tasks. Goddu et al.'s case is thus not sensitive to empirical considerations of model performance, regardless of how successful or unsuccessful that performance might be. Their argument is not that LLMs fall short of certain cognitive skills; rather, they maintain that LLMs' performance on tasks measuring cognitive skills is irrelevant on a priori grounds. If LLMs are not biological organisms, then they are not appropriate candidates for epistemic agency.

Yildirim and Paul´s (2024a, 2024b) claims, however, do not make any reference to LLMs' status as organic agents. For them, LLMs can also exhibit instrumental knowledge on a functional basis. Clearly, the variant of instrumental knowledge possessed by LLMs does not rely on, or suffice for, the enactivist criteria of organic agency. For Yildrim and Paul, however, LLMs raise the possibility of "novel" types of instrumental knowledge, pushing us to reconsider the traditional enactivist formulation of epistemic agency.

We contend that this debate hinges on how instrumental knowledge should be understood. If instrumental knowledge is interpreted in a broader sense to only mean the capacity to use some instrument (such as the linguistic input in the next-word prediction task), assigning it to LLMs becomes trivial. If, on the other hand, a more epistemically demanding reading is required, this kind of instrumental knowledge goes beyond the capacity to use the states of LLMs as instruments. Traditionally, such knowledge is taken to require the system under discussion to be an epistemic agent that has beliefs, or cognitive states (Williamson, 2000; Nagel, 2013). This raises a fundamental question: which systems deserve such a cognitivist interpretation in the first place?

In sum, this comes down to philosophical intuitions concerning agency, instrumental knowledge, and epistemic systems. Yildirim and Paul clearly maintain a different notion of agency than the enactivist one maintained by Goddu et al., which renders the latter's objection misplaced. At the same time, instrumental knowledge – as understood in the traditional epistemological literature – is assigned only to epistemic agents to begin with. In the absence of an independent motivation to treat LLMs as introducing a novel type of epistemic agency, their success at using an instrument (such as next-word prediction) remains insufficient to justify claims that they have attained knowledge, instrumental or otherwise. The resolution to the present predicament thus stems not from disagreements concerning LLMs as such, but philosophical intuitions concerning epistemic agency and knowledge.

Louis Loock

1.3 - SMART MINDS AND SMARTER TOOLS: HUMAN COGNITION IN THE AGE OF AI

Louis Loock, Osnabrück University

Abstract:

1 – Facing Intelligent Tools

What is the future of human cognition in a world of much more intelligent tools? Current efforts of engineering ever more capable digital tools mainly evoke philosophical concerns directly about the nature and ethics of artificial intelligence. But it might be more relevant to first ask how the daily usage of AI tools could impact our own cognitive abilities, and what this would reveal about the nature and ethics of our own natural intelligence. Prioritizing this question seems advisable, also because it could consequently redefine our immediate views on AI, too.

The present investigation from the field of situated cognition (Newen et al., 2018; Robbins & Aydede, 2008) advances a new perspective on our cognitive relations with external tools. Specifically, the very influential sub-debate on extended cognition (Clark, 1997; Clark & Chalmers, 1998) exclusively viewed our tool relations as cooperative and internally beneficial (Clark, 2008). But what if our current tool practices are rather detrimental for the standing of our internal cognition, especially if some form of cognitive replacement ensues (Gerlich, 2025; León-Domínguez, 2024; Paglieri, 2024)? Extended cognition is inherently unable to capture this possibility (Aagaard, 2021). Hence, another situated contender might be called for.

Consider how many of us are nowadays personally inclined, and perhaps even structurally incentivized, to utilize digital tools that can solve our cognitive tasks for us – or at least produce comparable outputs. This is realized via advanced technologies that can somewhat obtain parts of our cognitive skills which we would usually exert internally for those tasks. Our interaction strategies with those tools may then slowly decrease our own cognitive engagements and responsibilities. This might ultimately render us “extracted cognizers”. The hypothesis of extracted cognition states that we naturally, or at least habitually, desire external tools that solve our cognitive tasks independent of us, namely by making or letting them capture, mimic, and eventually replace those of our cognitive skills we would otherwise employ and train internally.

Three questions shall lead us to this hypothesis: First, how do we make intelligent tools? Second, how do we use intelligent tools? And third, how do we thereby become extracted cognizers?

2 – Making Intelligent Tools

For any tool we create, we need some template. Though to make tools intelligent, we must take ourselves as the template. Three general techniques are available.

First, there is designed extraction. For example, building a calculator or some expert system requires studying some type of human knowledge and reasoning, so to then devise a general logical structure from it. Though manually designing machines with more general abilities quickly reached a practical end with symbolic AI (Dennett, 1984).

A second, more ingenious technique is collective extraction as used by subsymbolic AIs. Deep learning applications require tons of pre-classified training data to enable their statistical feats. But instead of hand-crafting these sets, vast digital infrastructures scatter this data demand into millions of micro-tasks that are then solved by large collectives (Mühlhoff, 2020). Alternatively, if collectives already provided some template, machine learning techniques (like LLMs) can capture and mimic the cognitive skills originally employed for it (e.g., texts, images, videos, audio, or code).
A third technique of individual extraction gains traction just now. The fine-tuning of model parameters can be done with rather little individual data, and then those models can emit outputs just like their individual user. For example, an LLM can be fed with the oeuvre of Daniel Dennett to make it produce philosophical texts of indistinguishable style (Schwitzgebel et al., 2024).
Now, making intelligent tools surely requires our intelligence as some kind of template, but this does not yet illuminate how we then use those tools.

3 – Using Intelligent Tools

We use smart tools in two basic ways: Either we actively provide them with our own skills, or others already built them for desired tasks. The first tactic, cognitive outsourcing, occurs for example when one uploads their face, voice and a script to an AI that then fakes their virtual presence in online meetings (see HeyGen©). The second tactic, cognitive insourcing, might be to download an app that can read, summarize, and answer our questions on given texts (see NotebookLM©).
These neutral descriptors alone are however undecisive. A general economic principle guides our practices (Clark, 2007; Rowlands, 2010): 'Use the most lucrative resource for your cognitive task wherever its location.' Combining all this may reveal that we mostly let tools solve our tasks without us – against the conception of extended cognition.

4 – Becoming Extracted Cognizers

Extended cognizers strive for more and deeper engagements that result in cognitive integration with their tool (Menary, 2007). Extracted cognizers aspire the opposite result, because the tool performance cancels any of their involvements (Heinrichs,2024). Now, independent sets of theoretical, practical, ethical, and scientific criteria can each delimit extracted from extended cognition (cf. references).
Theoretical criteria address our theory of cognition. First, the degree of causal control executed by the agent over the tool is very low for extracted cognition (Piredda,2017). Second, the degree of epistemic access of the agent to the tool is also very low in extracted interactions (Andrada et al., 2023). Third, the decisional power in a cognitive task is hardly exercised by an extracted cognizer, but by their smart tool (Clark, 2007).
Practical criteria address our cognitive practices (Clark, 2003). First, extracted agents have the intention to let tools solve their tasks, and then stop thinking and learning themselves. Second, the development of such practices renders the internal skills of the agent obsolete or overly specialized. Ethical criteria address our ethics of cognition (Carter & Palermos, 2016). First, the agent's role in the interaction is massively weakened to a mere adjunct. Second, the tool function is to realize this cognitive substitution and replacement. Third, if our axiology hence judges this interaction as unfavorable, then we shall ascribe extracted cognition to the agent.

Scientific criteria address our science of cognition (Rupert, 2009). First, our methodology would aim to explain how exactly such technologies can take over ourcognitive abilities, tasks, and responsibilities. Second, our explanations of cognizers and their tools would describe them as two interacting, yet independent systems. Third, our conception of cognition would understand its nature as transferrable to, and replaceable by, certain external vehicles and their specific usage.

5 – Situating Extracted Cognizers

The provided answers form three crucial steps toward the future of human cognition. First, making intelligent tools requires them to capture our cognitive skills via abstract templates that we provide to them in different forms. Second, using intelligent tools is often driven by our tendency toward the most economic task strategy, and this often results in situations where our tools shall just imitate our cognitive skills that we previously provided to them. Third, becoming an extracted cognizer requires that those tools ultimately replace our cognitive skills as we no longer exercise them ourselves, and this result can be characterized by independent criteria from the theory, practice, ethics, and science of cognition. If these three answers are convincing, the future of human cognition might indeed best be described by the hypothesis of extracted cognition: We seek tools that solve our cognitive tasks like us, for us, and yet without us.

Yizhi Li

2.1 - WHY IS MIND WANDERING EFFORTLESS?

Yizhi Li, Ruhr-University Bochum, Institute for Philosophy II

Abstract:
When people rest on a sofa, take a shower, or stroll through a park without specific task demands, their minds remain active, often engaging in spontaneous, self-generated thought (mind wandering). Despite this persistent mental activity, such moments are widely perceived as relaxing. Beyond physical rest, a necessary factor contributing to this relaxation is the phenomenological lack of effort associated with mind wandering (MW). This low-effort quality is not merely incidental; it is an important feature of MW. For instance, Stan and Christoff (2018, p. 47) conceptualize MW as the mind "moving with ease," emphasizing its "absence of difficulty or effort". However, a theoretical paradox arises when we consider influential accounts of MW that appeal to cognitive control (Shepherd, 2019; Smallwood & Schooler, 2006, 2015). Cognitive control— the capacity to coordinate mental resources toward goal-directed behavior (Cohen, 2017)—is typically associated with mental effort. If MW requires cognitive control, why is it not experienced as effortful? This paper resolves this paradox by integrating the often-overlooked segmented structure of MW with the opportunity cost theory of mental effort (Kurzban, 2016; Kurzban et al., 2013). I argue that MW's low-effort quality stems from its rapid, frequent shifts between goal-relevant topics (segments), which prevent the accumulation of opportunity costs (and thus mental effort). This account reconciles the cognitive control view of MW with the low- effort phenomenology of MW. In the following, I will first outline the concept of mental effort and the segmented structure of mind wandering. Next, I will articulate the paradox inherent to the cognitive control view of MW and propose a resolution.

Ulrich Krohs

2.2 - A STATE MODEL OF CONSCIOUSNESS: FROM ELEMENTARY QUALITIES TO INTENTIONALITY

Ulrich Krohs, Department of philosophy, University of Münster

Abstract:
Some neural systems, such as subsystems of human brains, do have states that are conscious states. Other of their states, such as inactive states, are probably not conscious, and dead states certainly are not. Since neural activity is generally transient, conscious states are unstable and have a certain mean life. They arise and decay, as any state of a physiological system. The paper combines these simple considerations with the assumption that complex conscious states are fusions of simpler conscious states (Humphreys 1997). This combination provides the basis of a powerful model of conscious processes. The model explains, on a principled level, not only the dynamics of conscious processes, but also the generation of intentionality. It also suggests a possible mechanism for mental causation in a physical world, which cannot be explored in depth in this paper. In developing the model, I assume that the simplest conscious states are states of not too complex neural systems, far below the level of whole brain regions. They could be certain active states of neural networks, of systems of only a few neurons, or even of single neurons. The model considers these simplest, "micro-conscious" states as being independent consciousnesses. They are generated by excitation of the ground state of the neural system and vanish by spontaneous relaxation.
We can describe the relaxation in terms of mean life, which can be taken as an intrinsic property of the conscious state. Being conscious states, they have another intrinsic property, namely a conscious quality. While the mean lifes are measurable in principle, we do not have a clue as to how to assess the qualities. But we can assume that, if there are different types of conscious states, they differ in quality as well as in mean life. So one can tell the qualities by the mean life. Two or more systems in their activated state can merge. The independent consciousnesses fuse, thereby losing their independence. They assume a shared state and form one single consciousness. Paul Humphreys has drawn the analogy to quantum mechanical entanglement (Humphreys 1997), where the entangled systems also share a state. The entangled state variable can no longer be attributed to any of the components, but only to the entangled system. In the case of activated neural systems, the equivalent of an entangled state might simply be a state of synchronized activity. I call this kind of fusion 'macroscopic entanglement' to emphasize the aspect that complex consciousness is a unique or entangled state of the combined system. This metaphor is not meant to be related to any idea of quantum consciousness. Macroscopically entangled conscious states are unstable like any conscious state, so they can separate. This is, of course, only the simplest case of a complex consciousness. Strictly speaking, the contents in this case are not yet intentional objects, but qualia. But they are had by a complex consciousness and no longer isolated qualities. We would have to reconstruct the fusion of qualia to intentional objects, the relation of qualia and of intentional objects to the external world, and many other steps in order to reconstruct the consciousness of an individual from these simple elements. In my paper, I can only give some hints on how this can be worked out. Instead, in the remainder of my talk, I will focus on the dynamics of conscious processes. The stability or mean life of the entangled state, like the stability or mean life of independent micro-conscious states, is an intrinsic

property of the state that will somehow correspond to the qualities that make up the content of the complex consciousness. Thus, the dynamics of complex conscious systems depends on their content. Therefore, thinking, judging and acting depend on and ‘handle’ the content of mental states.

Tobias Schlicht

2.3 - Assessment or Attribution of Consciousness in AI Systems?

Tobias Schlicht, Ruhr-University Bochum

Abstract:
Butlin et al (2023) outline a research program for assessing consciousness in AI. They assume computational functionalism as a working hypothesis since it allows for conscious AI in principle and claim that well supported neuroscientific theories of consciousness can help us to assess consciousness in AI. First, I will show that computationalism and functionalism are two diCerent assumptions with more commitments than recognized in their research program. To get functionalism from computationalism, we also need an additional assumption, such as that the nature of mental states is (entirely) computational. To get computationalism from functionalism, we also need the independent assumption that all functional states are computational (Piccinini 2020). Neither view is especially plausible, especially regarding consciousness. Second, I will argue that we face significant epistemic limitations once we follow the second part of the proposal, relying on neuroscientific theories. Using Global Neuronal Workspace Theory (GNWT) as a popular example, I argue that while the four alleged signatures of consciousness may be empirically well supported, the derived computational model relevant for the assessment of consciousness in AI is not. Computationalism requires not only multiple realizability, but medium independence (Haugeland 1997) which is a stronger demand. A function is medium independent if it does not put any constraints on the physical properties of the realizing mechanism. Chess is an example of medium independence, while sports like snooker or football are medium dependent since their performance depends on physical properties of the medium in which they are played. While digital computation may be medium independent, neural computation may be not. And this may be a reason for Piccinini (2020) to suggest that neural computation is sui generis, diCerent from digital and analog computation. Relying on prior work by Chirimuuta (2022), Cao (2022) and Block (2005), I make a case for the medium dependence of neural processing. Part of the argument appeals to the unique role and features of chemical synapses (Pereda 2014) and their interactions with electrical synapses for information processing across neurons. These may be crucial for the neuronal implementation of a workspace in human brains. The point is that the appeal to GNWT (or any neuroscientific contender) does not put us in the epistemic position to decide whether the biochemical details of neural computation matter for the function of global broadcasting to yield consciousness. Thus, we cannot decide in favor of computational functionalism against a biologist competitor. In the absence of neurophysiological markers when we aim to assess whether an AI system is conscious, and the unreliability of a significant number of alternative markers of consciousness (Bayne et al. 2024), such as report (the outputs of Large Language Models do not persuade us that they are conscious), we are left with the attribution of consciousness from the intentional stance (Dennett 1987) rather than the possibility of an assessment of consciousness in AI systems. Therefore, the outlined research program faces significant limitations.

Benjamin Peters & Asya Achimova

3.1 - INFERRING TRUTH: A PSYCHOLINGUISTIC LOOK INTO DIRECT AND ACCOMMODATED EVIDENCE

Benjamin Peters, Osnabrück University

Sila Sevi Capar, Osnabrück University

Peter König, Osnabrück University, Department of Neurophysiology and Pathophysiology, Center of Experimental Medicine, University Medical Center Hamburg-Eppendorf

Asya Achimova, University of Tübingen

Abstract:

Presuppositions are a class of inferences that can be drawn from utterances. Take the following example:

(1) Emma stopped eating meat.

This utterance allows us to reach two conclusions:

(2) Emma currently does not eat meat.

(3) Emma used to eat meat.

While (2) represents a regular entailment, in the literature conclusion (3) is referred to as a pre- supposition. In contrast to entailments, the content of presuppositions is generally assumed to be shared knowledge between the interlocutors at the time of the utterance (Stalnaker, 1973; Karttunen, 1974). If the relevant content is not already part of the common ground of the interlocutors, the speaker implicitly asks the other discourse participants to accommodate the information, i.e., adding it to the common ground or at least accepting that the speaker is committed to that information for the purposes of the current interaction (Lewis, 1979; Beaver and Zeevat, 2007). In this case, we speak of an informative presupposition (Stalnaker, 2002; Von Fintel, 2008). Presuppositions can also be distinguished from entailments in terms of their behaviour in complex sentences. When embedding the initial example above under entailment-cancelling operators like negations or conditional constructions, we can still draw the inference in (3), but not the one in (2). In linguistics, this is called projection. Due to their implicit nature, presuppositions are frequently used as a tool for persuasive com- munication in domains such as politics or advertising (Lombardi Vallauri, 2021):

(4) We will make America great again.

(5) Stop watching, start living!

Presuppositions can arise from a variety of linguistic constructions, called presupposition triggers. This set of expressions is heterogeneous in several aspects. For example, prior research has pointed out differences between and even within classes of presupposition triggers regarding their evoked projection inferences (Kiparsky and Kiparsky, 1970; Simons et al., 2017), i.e., conclusions about whether a speaker is committed to the truth of the embedded content (Tonhauser et al., 2018; Degen and Tonhauser, 2022). Prior beliefs towards the subject of the presupposed information further influence projection behaviour (Mahler, 2020; Degen and Tonhauser, 2021), meaning that less plausible content projects less. Meanwhile, Mazzarella et al. (2018) provided empirical evidence that presuppositions are

taken to be equally committal as assertions. Here, we focus on investigating whether the choice of trigger type influences information acceptance— the extent to which participants find the presupposed information believable. We further explore how plausible presupposed content is in contrast to asserted content by comparing their respective information acceptance degrees. Since informative presuppositions are occasionally used to embed disputable information, e.g., in political discourse (Masia, 2020), we particularly created items con- taining falsehoods about historical events and investigated how certain participants were that the false information was true. Furthermore, we examined whether (self-reported) prior knowledge in- fluences information acceptance.

Amit Singh

3.2 - EMBEDDING PSYCHOLINGUISTICS: AN INTERACTIVE FRAMEWORK FOR STUDYING LANGUAGE IN ACTION

Amit Singh, Psycholinguistics, Faculty of Arts and Humanities, Paderborn University

Katharina J. Rohlfing, Psycholinguistics, Faculty of Arts and Humanities, Paderborn University

Abstract:
There has now been substantial evidence accumulated in the favour of language, vision, and action processes dynamically and continuously interacting with one another in a bidirectional manner within a richly embedded context (Spivey, 2007). Building on this foundation, numerous studies have demonstrated that linguistic processes can influence visual context in real time, and conversely, that visual context can shape language comprehension the moment it is made available to the sensory stream (Altmann & Kamide, 2009). From the perspective of situated cognition, cognitive processes are assumed to be inherently grounded in real-world environments, intrinsically involving both perception and action (Wilson, 2002, p. 626).

These perspectives mark a departure from traditional modular approaches in cognitive sciences (Dingmanse et. al, 2023; Spivey, 2023), which often conceptualized mental processes as serial and functionally encapsulated, akin to the computer metaphor of the mind (Rescorla, 2015). Instead, advancing beyond these paradigms necessitates interactive setups and calls for a 'massive interactivity' approach — a framework that contrasts with previously dominant notion of 'massive modularity' over the past several decades (Fodor, 1983).

Whereas prior studies have significantly contributed to this paradigm shift by employing online methodologies to investigate real-time processing, fully appreciating the interactivity — and capturing the bidirectional influences among language, vision, and action within ecological contexts — requires experimental setups that emulate naturalistic, everyday interactions. However, a common critique of such situated designs is often highlighted by the potential lack of experimental control over the key latent variables. To address this challenge, the present study introduces a novel methodological framework that integrates psycholinguistic paradigms into a Human-Robot Interaction (HRI) setup. Our setup enabled controlled investigation of the interdependencies among language, vision, and action in a situated environment. We first present results examining how language and action conceptualization influence visual attention using an eye tracker and then discuss how this setup might offer a promising avenue for studying real-time, bidirectional interactivity among these domains, comparable to experimental rigor to classical psycholinguistic paradigms, such as on-screen visual world paradigms (Tanenhaus et. al., 1995, VWP).

Our study was motivated by one of the most frequent and ubiquitous phenomena in real-world interaction: conversational breakdowns and interactive repair (Albert & de Ruiter, 2018). Interactive repairs standing at the heart of interaction, they are suggested to not only dependent on the immediate conversational context but also reflects interlocutor's accumulated knowledge regarding the source of misunderstanding (Dingmanse and Enfield,

2015, pp 105-106). More specifically, thus, the conceptualization of the misunderstanding—manifested through different forms of repair—is predicted to modulate attentional processes (Dingmanse & Enfield, 2023). For instance, it has been theoretically proposed that distinct types of repair encodings have differential consequences for attentional allocation. Open or general repairs (e.g., "How?" or "Where?") do not specify the locus of misunderstanding and therefore require less attention to the context. In contrast, restricted repairs (e.g., "Did you mean putting it vertically or horizontally?") explicitly address the specific source of the misunderstanding, demanding more focused attention on the unfolding situational context, or the simulation of actions if the interaction involves sequential motor activities. If the assumptions about the interactive nature of cognitive processes are consistent with findings from previous psycholinguistic studies, then the linguistic encoding of repairs—as well as the mental simulation of actions—should be observable in participants' visual behavior within a context.

Thus, to investigate the interplay between language, action, and vision during conversational repair, we designed an experimental setup in which a humanoid robot instructed participants to perform a sequence of actions using four distinct coloured objects placed in front of them. Each instruction specified both the manner of action (e.g., vertically or horizontally) and the path (designated by letters). Crucially, this setup adopted a trial-by-trial, repeated-measures design commonly used in psycholinguistics, allowing for controlled manipulation of critical trials where action information—either manner or path—was occasionally omitted. Importantly, the setup preserved interactivity by allowing participants to initiate repair (i.e., ask clarifying questions) when they perceived that necessary information was missing. Instructions followed a simple linguistic template referencing both manner and path, for example:

> “Lege bitte das rote Objekt hochkant auf A”
> (Translated: “Put the [COLOR] object [MANNER] on [PATH]”).

A sequence of actions, each associated with one of the objects, established a dynamic context upon which subsequent instructions were provided, and actions were performed. Critical trials with missing information were embedded within this sequence, prompting participants to engage in repair behaviour. To investigate contextual influence on visual attention, we manipulated the informational richness of the action sequences, creating two types of contexts: Simple and Rich.

In the Simple model, participants experienced only one type of manner (e.g., all actions were vertical). In contrast, the Rich model included both manners in the preceding actions, offering a more varied context. If the form of repair (open vs. restricted) influences mental simulation of action—as suggested by prior psycholinguistic research—we expected such effects to be reflected in participants’ visual attention patterns. Crucially, we hypothesized that these effects would differ depending on the contextual model (Simple vs. Rich), providing insight into how contextual variation modulate attention during repair encoding. We analyzed participants’ dwell times on the action models (Simple vs. Rich) as a function of repair type (Open vs. Restricted) in the critical trials. Overall, we observed a significant main effect of repair type, with Restricted repairs eliciting longer dwell times on the model compared to

Open repairs (β = 52.46, 95% CI [45.37, 59.65]). This finding suggests that encoding Restricted repairs requires greater attentional engagement with the contextual information provided.

Critically, this pattern changed and was sensitive to the visual context in front of participants. In Rich contexts—where both action types were present—Restricted repair requests led to substantially increased dwell times compared to Simple contexts (β = 241.72, 95% CI [227.15, 256.45]). This indicates that the visual context modulated the attentional allocation during repair initiation. Specifically, when a more informative action model was visually available, participants were more likely to direct attention toward it during the formulation of a Restricted repair. This finding supported the hypothesis that visual context influences the linguistic encoding of repair, particularly when the repair demands precise reference to the action repertoire.

Our setup allowed participants to engage in spontaneous, real-time repair—offering a powerful window into how language and vision might interact in situated settings. Unlike traditional psycholinguistic experiments that often rely on static, screen-based tasks with minimal interactivity, our approach allowed to investigate language, vision, and action in a dynamic setting. Our results show that participants' gaze patterns are modulated by both, the kind of repair they conceptualize, and the richness of the visual context provided. This suggests that language processing is not only about internal cognition but rather shaped by interactivity and situatedness, a factor often missing in screen-based experiments in labs.

Leandra Bucher

3.3 - Cognitive Processing of Probability Information in Attributions of Justification and Knowledge

Leandra Bucher, University of Siegen

Paul Thorn, University of Düsseldorf

Abstract:
Probability and frequency information are often considered objective markers of uncertainty. However, empirical findings show that human reasoners tend to process them in context-sensitive ways. Contextual cues such as positive and negative frames, linguistic formulations, source credibility – to name just a few - influence how individuals interpret and respond to probability and frequency information. Tversky and Kahneman's (1981) pioneering work on framing effects demonstrated that individuals make systematically different choices depending on whether identical probability outcomes are framed as gains or losses. Such findings suggest that probabilistic reasoning is not only a matter of mathematical competence but is also subject to cognitive heuristics and biases. Our research extends this line of inquiry by examining how framing, linguistic structure, contextual information, and source characteristics influence the interpretation of probabilistic statements. In a series of experiments, we manipulated key dimensions of probability presentation. Participants were presented with identical probabilistic data embedded in varying linguistic frames (e.g., "95% chance of success" vs. "95% chance of failure"), different contextual scenarios (e.g., medical diagnosis vs. legal judgment), and with variations in the described source of the information (e.g., expert vs. test result). Additionally, we examined how participants evaluate the strength of beliefs based on probability and whether they consider certain levels of probability sufficient to justify the attribution of knowledge. Preliminary findings suggest that individuals' assessments of whether a belief is justified, or whether someone "knows" a fact based on probabilistic information, are strongly modulated by contextual framing. For example, a 95% probability may be judged sufficient to claim knowledge in a medical context but insufficient in a legal context, where even higher standards of certainty are expected. These results support the hypothesis that probability information does not function in isolation but interacts with normative expectations and domain-specific knowledge. Moreover, our data indicate that the perceived credibility of the source of probability information plays a significant role. Probabilities derived from expert opinion tend to be weighted more heavily than those attributed by impersonal test results, even when the numerical values are identical. This preference suggests that social and epistemic trust modulates probability processing—a phenomenon that has implications for both applied decision-making and epistemological theory.

Martin Butz

4.1 - Controlling Meta-Control: Deriving and minimizing the Costs and Benefits of Meta-Control

Max Mittenbühler, Cognitive Modeling, Department of Computer Science and Department of Psychology, Faculty of Science - University of Tuebingen

Martin Butz, Cognitive Modeling, Department of Computer Science and Department of Psychology, Faculty of Science - University of Tuebingen

Abstract:
Human cognition frequently requires balancing automatic and controlled processing (Evans and Stanovich, 2013). Controlled processes offer flexibility and goal-directed be- havior but require greater cognitive effort. On the other hand, automatic processes are effort-efficient but less adaptable. Theories of resource rationality propose that humans opti- mize this trade-off to maximize utility under cognitive con- straints (Gershman et al., 2015; Griffiths et al., 2015; Lewis et al., 2014; Ortega and Braun, 2013; Shenhav et al., 2017). We fully agree. An open challenge is, however, how to formalize this tradeoff between automatic and controlled processing and how to balance it. Here, we propose how this is done. Prior frameworks have attempted to balance automatic and controlled processing via the formalization of control signal costs or information-processing costs (Kool et al., 2017; Shenhav et al., 2013). They rely on hand-tuned trade- off parameters and do not address how control and meta- control costs may interact Here, we propose a unified, normative framework, which is based on the variational free energy principle Friston et al., 2015. In other words, we propose a mathematical princi- ple that is able to estimate allows, via active inference, the inference of situation-specific costs and benefits of invok- ing additional control processes for decision making. The objective is derived from a formalization of surprise mini- mization in a (Bayesian) graphical model, which encodes a task-specific interaction event trial Butz, 2016 as a contextual prior, which can be further parameterized by density- and transition-manipulating parameterizations Butz et al., 2024 The inference process is implemented as a sampling process that estimates posteriors over potential actions and action- outcomes (e.g., predicting if pressing letter ’a’ would be a correct or incorrect response). We demonstrate that the sampling-based minimization of the variational free energy estimate simultaneously governs both the resource allocation between automatic and controlled processes and the meta-control mechanism itself. Exemplarily, we show a validation of the Stroop task (pre- viously published in Mittenbühler et al., 2024), where key behavioral effects, specifically the proportion congruency ef- fect (Spinelli and Lupker, 2023), emerge naturally. The re- sults, as well as further investigations, suggest that the sam- pling process approximates the optimal posterior under the constraint of minimizing surprise. Further, actual (sampling- based) minimization steps can be quantified as cognitive ef- fort. For example, in action space, which we focus on here, the effort of choosing an action can be quantified as a change in action execution tendency. Post-hoc, the effort can be quantified by the KL-divergence between the density before a stimulus was presented with the density when the action is actually executed after stimulus presentation. We can show that action executions (e.g., motor actions) should by invoked once the sampling process ceases to minimize surprise any further.

Viktoria Zemliak

4.2 - Familiarity Memory in Recurrent Spiking Neural Networks

Viktoria Zemliak, Osnabrück University

Pascal Nieters, Osnabrück University

Gordon Pipa, Osnabrück University, Frankfurt Institute for Advanced Studies

Abstract:
Familiarity − the sense that a stimulus has been encountered before − is a simple, yet efficient form of memory. Unlike semantic memory, which encodes the content of a stimulus, familiarity functions as meta-information: it signals prior exposure without requiring the recall of specific contextual or semantic details. Familiarity information can be beneficial for both artificial and living agents. For example, in continual learning tasks, familiarity estimation enables the reuse of existing representations for new tasks, enhancing performance and energy efficiency in spiking networks (Han et al., 2024). From a Bayesian reinforcement learning perspective (Strens, 2000), familiarity helps resolve the exploration–exploitation trade-off: novel stimuli promote exploration, while familiar ones support efficient exploitation. In this way, familiarity regulates adaptive behavior under uncertainty.

Neural correlates. Neural signatures of familiarity have been observed throughout the sensory processing hierarchy. In its simplest form, familiarity can reflect recent repetition and appear as rapid neural adaptation, resulting in repetition suppression (see Grill-Spector, Henson & Martin, 2006, for review). Interestingly, this suppression can sometimes persist for hours or even days. It has been reported in areas such as V2 (Huang et al., 2018), inferior temporal cortex, perirhinal cortex (Miller et al., 1991; Anderson et al., 2008; Meyer & Rust, 2018), and prefrontal cortex (Rainer & Miller, 2000). In contrast, several studies report stimulus-selective response potentiation in V1 for familiar stimuli (Cooke et al., 2015; Hayden et al., 2023). While the hippocampus is traditionally linked to episodic recall (see Montaldi & Mayes, 2010 for the discussion), Rutishauser et al. (2006) demonstrated that distinct hippocampal neuron types can signal familiarity through both enhanced and suppressed activity.

Mechanisms. Familiarity encoding operates across multiple levels and timescales. On a systems level, predictive coding frameworks propose that reduced responses to expected stimuli arise from top-down predictions suppressing redundant input (Friston, 2005; Summerfield et al., 2008). Repetition suppression may also result from enhanced inhibitory recruitment, supporting fast-scale adaptation (Kohn & Movshon, 2004; King et al., 2013). At the synaptic level, anti-Hebbian plasticity can suppress familiar input responses (Bogacz & Brown, 2013; Tyulmankov et al., 2022). Here, we explore an alternative manifestation of familiarity: Hebbian strengthening of recurrent connections through repeated exposure. For example, V1 neurons with similar receptive fields form stronger recurrent connections with each other (Cossel et al., 2015), which are shaped by acquired visual experience in early development via Hebbian plasticity (Schmidt et al., 1999; Sadeh et al., 2015). But how can familiarity be read out from the activity of a recurrent network? In this work, we compare two

decoding strategies: based on firing rate and on temporal coding, namely spike synchrony, which was shown to reflect underlying recurrent connectivity (Korndörfer et al., 2017). We show how both methods can decode familiarity from activity of a recurrent spiking network, shaped by unsupervised Hebbian learning, in dynamic environments, and identify the conditions under which one method outperforms the other, as well as when they need to cooperate.

Nick Augustat

4.3 - MAPPING REPRESENTATIONAL AND COMPUTATIONAL DYNAMICS DURING FEEDBACK PROCESSING IN A REINFORCEMENT LEARNING TASK

Nick Augustat, Department of Psychology, Marburg University

Li-Ching Chuang, Department of Psychology, Marburg University

Philipp Bierwirth, Department of Psychology, Marburg University

Erik Malte Mueller, Department of Psychology, Marburg University, Center for Mind Brain and Behavior, Marburg University and University of Giessen and Technical University of Darmstadt

Dominik Endres, Department of Psychology, Marburg University, Center for Mind Brain and Behavior, Marburg University and University of Giessen and Technical University of Darmstadt

Abstract:
Reinforcement learning (RL) guides behaviour by integrating rewarding and non-rewarding feedback into a successful action policy, presumably involving mesolimbic dopamine signalling. Electroencephalography (EEG) signals reveal different neural signatures in response to reward, punishment or neutral feedback, which are often assessed using isolated measures, such as the event-related potential (ERP) component amplitudes P300 and feedback-related negativity or frontomedial oscillatory power in the theta frequency band. However, the temporal interplay between attention, learning and working memory in RL all of which are intertwined as part of a cognitive control loop, may not be fully captured by any single ERP measure.
Our first aim was therefore to investigate how the temporal profile of feedback processing may differ between feedback conditions in a probabilistic selection task, and if it is affected by placebo expectancies and a dopamine pharmaco-challenge. To this end, representational similarity analysis (RSA) enables analysing time-resolved stimulus representations by correlating stimulus-locked patterns of oscillatory power across time, offering insight into how stable stimulus representations are over time. Moreover, dynamic coding describes that stimulus representations change over time, such that the pattern of oscillatory power associated with a given stimulus differs across time points: the higher the similarity within each time sample, compared to the similarity between time samples, the stronger the representational dynamicity. It is thought to support flexible processing by minimising interference between temporally adjacent events.
Our second aim was to probe if representational dynamicity during feedback processing can be mapped onto a computationally plausible RL model. Specifically, we speculated that higher representational dynamicity would indicate stronger integration of feedback information, similar to effects observed in long-term memory formation, where greater temporal specificity of neural representations has been associated with enhanced encoding and retrieval. In context of RL, such dynamic coding would allow the system to transiently separate feedback signals across trials, facilitating integration without interference. Computational RL models enable capturing sensitivity to recent feedback by means of learning rates (i.e., how strong a single outcome affects future reward expectations), optionally controlled for possible task-irrelevant confounders. The three-parameter-model assessed in this study resulted from

our extensive comparison among 74 candidate models (unpublished) ensuring the best fit to data given parameter recoverability.

Klaus Von Heusinger

5.1 - CAUSAL CONNECTIVES AND COGNITIVE FLEXIBILITY IN FIRST-EPISODE OF PSYCHOSIS

Klaus Von Heusinger, University of Cologne

Derya Cokal, University of Cologne

Emre Bora, Dokuz Eylül University

Abstract:

Background. Using connectives such as because mirrors cognitive tasks during speech as we plan what we will say next and establish relations between utterances to build coherence in discourse structures. Therefore, cognitive flexibility is a critical component of effective communication, enabling individuals to develop and elaborate a complex topic narrative structure during speech. Impairments in cognitive flexibility can interfere with this process, leading to fragmented or disrupted speech, particularly in patients with psychosis (Liddle, 1987; Little et al., 2019; Palaniyappan, 2021). The early stages of psychotic disorders, such as First- Episode Psychosis (FEP) and Clinical High-Risk Psychosis (CHR), provide a rich context in which to explore how cognitive flexibility deficits manifest in language use (Dajani & Uddin, 2015; Mackinley et al., 2021). Individuals with these conditions may struggle when transitioning between ideas and maintaining coherent discourse (e.g. Mackinley et al., 2023). In addition, Mackinley et al., (2023), found decreased use of overt connectives among English FEP and CHR groups. This difficulty in organizing thoughts and speech can be linked to cognitive inflexibility, which impairs the ability to adapt one's narrative structure according to conversational demands. In this research context, our study investigates the relationship between cognitive flexibility, coherence, and speech. In particular, we investigated the use of such causal connectives by FEP and CHR individuals, highlighting the role of cognitive flexibility in narrative coherence and the potential for early identification of cognitive dysfunction through linguistic analysis.

Participants. The sample included (n = 53) FEP participants, (n = 64) CHR individuals, and (n = 34) neurotypical controls (NCs). The FEP and CHR groups had a mean age of 21 (SD = 4.49) and 11.29 years of education (SD = 2.78), while the NC group had a mean age of 22.33 years (SD = 4.25) and 14.44 years of education (SD = 2.51). All subjects were native speakers of Turkish. While all FEP participants had a history of a first psychotic episode within the last 12 months, they were all currently "clinically stable", meaning that they were currently not having acute psychotic, manic, or depressive episodes and had not changed their medication in the last 4 weeks due to symptom exacerbation. The FEP participants underwent interviews utilizing the Structured Clinical Interview for DSM-IV Axis I Disorders (First et al., 2002). The Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen, 1986) and the Brief Negative Symptom Scale (BNSS) (Weigel et al., 2023) were used to assess current positive and negative symptoms in patients. The patients' current social functioning was evaluated using the Personal Social Performance (PSP) (Morosini et al., 2000). CHR participants met the criteria for one of three prodromal syndromes, as assessed using the Structured Interview for Psychosis-Risk Syndromes/Scale of Psychosis-Risk Symptoms (SIPS/SOPS: Miller et al., 2003; Tonyalı et al., 2022): (1) Delusions or hallucinations that subside over time); (2) Early symptoms are less

severe than fully developed psychosis; and (3) Family history of psychosis may lead to a decline in social, occupational, or academic functioning.

Task. Each participant had one minute to describe eight images from the Thought and Language Index (Liddle et al., 2002). Participants' responses were transcribed using the F4 program. Codings were agreed upon by two native speakers of Turkish.

Annotation. We annotated all causal discourse relations in the descriptions. We annotated causal discourse relations as implicit (without an overt marker as in (e.g., "I feel tired. I didn't sleep well.") or overt (with an overt marker as in (e.g., "I feel tired, because I didn't sleep well."). We focused on causal relations as only one study (i.e. only with English individuals) has shown overt causal connectives can be a good indicator of thought disturbances in psychotic speech 2 (e.g. Mackinley et al., 2023). Additionally, they can serve as a critical biomarker to detect speech deficits and cognitive inflexibility to allow for early intervention. Furthermore, following Çokal et al.'s (2020) and Sanders & Spooren's (2015) integrative approach to subjectivity, we tagged a causal relation as objective or subjective. A causal relation was annotated as objective if it mentioned a factual statement (e.g., The temperature rose because the sun was shining.), and subjective if it referred to a narrator/speaker's personal beliefs or interpretation (e.g., I think the man will attack her because he is stalking her.).

Predictions. We predicted FEP and CHR groups would demonstrate significantly reduced use of causal connectives compared to the NC group. This reduction in the use of causal connectives is likely related to impaired cognitive flexibility, as connectives help in organizing thoughts and transitioning smoothly between ideas. The observation of a reduced frequency of causal connectives could help in identifying cognitive-linguistic impairments before they fully manifest in clinical symptoms. Further, we explored whether the FEP and CHR groups would demonstrate more subjective causal reasoning (e.g., relating events to personal beliefs or delusions) compared to controls. Data analysis. Negative binomial regression (with utterance count as offset) tested group differences in causal relation frequency, controlling for age and education. P-values were FDR corrected (reported as q-values). A Kendall's Tau test examined correlations with clinical symptoms (thought disorder, negative/positive symptoms, and scores from BNSS, SAPS, and PSP).

Results. With respect to each group, we see that the NC group used a balanced number of overt and implicit causal relations (NC: implicit: 53%; overt: 47%), whereas both FEP and CHR groups used many more implicit causal relations and less overt connectives (see Figure 1; FEP: implicit: 73%; overt: 27%; CHR: implicit: 66%; overt: 34%; FEP: overt uses: $\beta = -0.772$, $z = -3.255$, $p = .001$; CHR: overt uses: $\beta = -0.627$, $z = -2.657$, $p = .008$). Both FEP and CHR groups used more subjective causal relations than the NC group (See Figure 2). Notably, the FEP group produced significantly more subjective causal relations than the NC group (NC vs. FEP: $\beta = 0.964$, $z = 3.397$, $p = .002$). The CHR group did not show a significant difference from the NC group in the use of objective and subjective causal relations (all p's $> .005$).

Derya Cokal

5.2 - Speech Disfluency as a Cognitive-Linguistic Marker of Early Psychosis

Derya Cokal, University of Cologne, Institute for German Language and Literature I – Linguistics University of Cologne

Klaus Von Heusinger, University of Cologne, Institute for German Language and Literature I – Linguistics University of Cologne

Emre Bora, Dokuz Eylül University, University of Melbourne and Melbourne Health

Abstract:

Background. Speech is inherently marked by pauses that reflect underlying cognitive processes and mental effort (Bello-Lepe et al., 2024; Çokal et al., 2019; Schilperoord & Sanders, 1997). Pauses between 250 to 3,000ms are considered natural components of typical speech and cognitive functioning (Goldman-Eisler, 1958; Schegloff et al., 1977), and interpreted as moments of cognitive delay or feedback (Goodwin, 1979) during which the semi-automatic flow of speech allows for upcoming content planning. Importantly, the syntactic location of pauses signals cognitive load: (a) utterance-initial pauses are linked to planning (Butterworth, 1979; Levelt, 1989), while (b) within-utterance pauses indicate lexical or syntactic processing (Hartsuiker & Notebaert, 2010; Kircher et al., 2004). Similarly, pauses (i.e., unfilled pauses) and fillers (e.g. erm, ah) are associated with speech monitoring (Levelt, 1983) and social signalling (Howes et al., 2017; Lake et al., 2011). In addition, several studies show the role of speech dysfluencies as biomarkers of mental health problems (Rapcan et al., 2010; Stanislawski et al., 2021). Since different types of disfluencies may serve different cognitive functions, such dysfluency patterns can provide a useful comparative window into cognitive dysfunction in psychosis (Çokal et al., 2019; Matsumoto et al., 2013; Parola et al., 2020) and can serve as an early signature of mental health problems (Parola et al., 2020; Stanislawski et al., 2021). However, most findings are based on studies of chronic psychosis patients, and there remains a lack of fine-grained differentiation of pause types based on syntactic position, as well as limited cognitive-theoretical grounding. Only one study has shown that chronic psychosis patients produce significantly more unfilled pauses than neurotypical controls (NCs) in utterance-initial positions and before embedded clauses (Çokal et al., 2019). To address this gap, we investigated whether specific speech disfluencies reflect cognitive processes and dysfunction in first-episode psychosis (FEP), whether the FEP group presents similar speech dysfluencies to those found in chronic psychosis patients (Çokal et al., 2019), and whether dysfluency patterns are associated with clinical symptoms. The FEP group is unique since it represents the early phase of psychosis, offering valuable insights into the cognitive and linguistic impairments that emerge before the effects of long-term treatment, thus contributing to our understanding of the cognitive and linguistic mechanisms underlying psychosis thought.

Participants. The sample included 53 FEP participants and 34 neurotypical controls (NCs). The FEP group had a mean age of 21 years (SD = 4.49) and 11.29 years of education (SD = 2.78), while the NC group had a mean age of 22.33 years (SD = 4.25) and 14.44 years of education (SD = 2.51). FEP participants had a history of a first psychotic episode within the last 12 months

and were all currently clinically stable. The criteria for clinical stability were to include individuals not having acute psychotic, manic, or depressive episodes and those who had not changed their medication in the last 4 weeks due to symptom exacerbation. All subjects were native speakers of Turkish. The participants underwent interviews utilizing the Structured Clinical Interview for DSM-IV Axis I Disorders (First et al., 2002). The Scale for the Assessment of Positive Symptoms (SAPS) (Andreasen, 1986) and the Brief Negative Symptom Scale (BNSS) (Weigel et al., 2023) were used to assess current positive and negative symptoms in patients. The patients' current social functioning was evaluated using the Personal Social Performance (PSP) (Morosini et al., 2000).

Task. Participants described eight images from the Thought and Language Index (Liddle et al., 2002) for one minute each (Çokal et al., 2022). Participants' responses were transcribed using the F4 program. Codings were agreed upon by two native speakers of Turkish.

Annotation. Disfluencies were annotated by syntactic position: utterance-initial pause (e.g. [PAUSE] He phoned his friend), within-clause pause (e.g., He phoned [PAUSE] his friend.) and 2 before-embedded clause pause (He phoned his friend [PAUSE] whom he has not seen since the graduation.) Duration of pauses (e.g., < 1 sec and between 1-3 secs) was annotated. As in previous studies, we excluded very short pauses less than 250ms, (Çokal et al., 2019; Howes et al., 2017; Matsumoto et al., 2013) and pauses longer than 3seconds (Kircher et al., 2004; Matsumoto et al., 2013; Rapcan et al., 2010). Pauses less than 250ms are phonatory gaps linked to the respiratory cycle, (Goldman-Eisler,1968), which are very hard to discriminate (Campione & Véronis, 2002). Pauses longer than 3seconds signal that a participant was no longer engaged with the given task. Additional annotations included filled pauses (e.g., [FILLER: er] He phoned his friend.), repetitions, and truncated utterances. Annotation followed prior protocols (Çokal et al., 2019).

Predictions. We predicted the FEP group would exhibit greater overall dysfluency—particularly at embedded clause and utterance boundaries—reflecting disruptions in higher-level speech planning. Given the association between fillers and social signalling, the FEP group would produce fewer fillers relative to the non-clinical groups.

Data analysis. Poisson regression models (with utterance count as offset) tested group differences in disfluency frequency, controlling for age and education. P-values were FDR corrected (reported as q-values). A Spearman rank permutation test examined correlations with clinical symptoms (thought disorder, negative/positive symptoms, and scores from BNSS, SAPS, and PSP).

Results. The number of utterances produced by NCs was 50 (SD = 13.86) and FEP was 41 (SD = 12.69). As shown in Figure 1, fillers were significantly reduced in FEP compared to NC, possibly reflecting reduced metacognitive monitoring or self-regulation in speech. Total pauses, including all syntactic positions of unfilled pauses, were highest in FEP ($\beta = 0.34$, $p < .001$). Utterance-initial pauses increased from NC to FEP, indicating potential speech planning and narrative difficulties (FEP: $\beta = 0.58$, $p < .001$). FEP participants also showed longer pauses (1–3 seconds). Within-clause pauses were more frequent in FEP ($\beta = 0.35$, $p < .001$), suggesting

lexical or syntactic difficulties related to cognitive load. Between-clause pauses were fewer in FEP (β = - 1.28), reflecting reduced syntactic complexity and less natural pausing. Repetitions were significantly higher in FEP (β = 1.68, p < .001), indicating disruptions in phonological loop functioning or self-monitoring. Truncated utterances, indicative of abandoned speech, were more frequent in FEP (β = 1.81, p < .001), consistent with thought derailments, planning failures, or difficulty maintaining communicative goals. Unlike Stanislawski et al. (2021), no disfluency measures were significantly associated with clinical symptoms and the scores from BNSS, SPAS, and PSP. Consistent with our predictions, the FEP group demonstrated distinct speech disfluency patterns compared to NCs, particularly in domains associated with reduced self-monitoring—evidenced by increased total pauses and fewer fillers (Parola et al., 2020). This reduction in pauses and fillers aligns with our hypothesis that social-signalling functions of speech may be diminished in early psychosis (Çokal et al., 2019; Matsumoto et al., 2013). The FEP group produced more within-utterance pauses, potentially reflecting lexical retrieval or syntactic integration difficulties. This finding is consistent with prior observations in chronic patients (Çokal et al., 2019). The lack of correlation with clinical symptoms reflect that dysfluencies capture cognitive processes that are not fully accounted for by traditional symptom scales. We propose that dysfluencies would be sensitive cognitive and linguistic loci of disruption in FEP and can serve as non- invasive, behaviourally grounded indicators of underlying cognitive disturbance.

Jannik Sühling

5.3 - DIFFERENCES IN COGNITIVE FOCUS ON MEANING VS. FORM IN AUDITORY LANGUAGE PROCESSING BETWEEN AUTISTIC AND NON-AUTISTIC ADULTS

Jannik Sühling, Ruhr-Universität Bochum

Eva Belke, Ruhr-Universität Bochum

Abstract:

Autism spectrum disorders (ASD) are neurodevelopmental conditions characterized by a wide range of symptoms, including variations in linguistic and communicative abilities. Recent studies have shifted away from rigid subtype classifications in favor of viewing autistic symptoms as existing on a continuum with high variability in language competence (Tebartz van Elst et al., 2021). In particular, while some individuals with ASD exhibit typical or even above-average language skills, others show marked linguistic impairments (Eigsti et al., 2011; Girolamo & Rice, 2022; Haser et al., 2021; Kjelgaard & Tager-Flusberg, 2001, Matsui et al., 2022). For those without overt language impairments, observed differences in language processing are often attributed to a distinct cognitive processing style rather than to a comorbid language disorder. Previous psycholinguistic research, primarily conducted with children, has highlighted that individuals with ASD tend to focus on lower-level, phonological information – a manifestation of a local processing bias (Happé & Frith, 2006). Yet, it remains an open question whether such detail-focused processing persists in adulthood. Drawing upon the seminal auditory language study by Goh et al. (2016), which demonstrated pronounced phonological neighborhood effects in lexical decision (LD) tasks as well as concreteness effects in semantic categorization (SC) among non-autistic adults, the present study aims to replicate and extend these findings in a German-speaking autistic sample. Specifically, we investigate two hypotheses in verbally unimpaired adults:

1. Autistic individuals will show stronger phonological neighborhood effects in the LD task than neurotypical participants.
2. Autistic individuals will exhibit weaker concreteness effects in the SC task compared to neurotypical controls.

Aïda Elamrani

6.1 - THREE LEVELS OF DISTINCTION BETWEEN BIOLOGICAL AND ARTIFICIAL CONSCIOUSNESS

Aïda Elamrani, Institut Jean Nicod, ENS-PSL & Chargée d'études CNRS

Abstract:
Consciousness is a complex phenomenon with manifest biological ties, which we analyse across three abstraction levels. At the physical level, a biological substrate, composed of cells and DNA, is often invoked as a plausible demarcation between conscious and non-conscious beings. At the phenomenal level, biological constraints such as embodiment, survival or reproduction profoundly shape psychology and subjective experience. At the cognitive level, the range of input-output relations computed by conscious processes are bound by biological algorithms such as feedforward neural networks or Hebbian learning. We argue that if the imitation or replication of either one of these levels is required for consciousness, then artificial systems remain far from achieving the kind of subjective states humans experience. As AI systems grow more sophisticated, distinct terminology may be necessary to differentiate advanced cognition from biological consciousness. While the scientific question of consciousness is unlikely to be resolved within the next decade, increasingly advanced artificial agents will proliferate, raising pressing public questions about their integration and our tendency to attribute consciousness to them.

Henry Shevlin

6.2 - DIGITAL MINDS? RECONCILING FOLK PSYCHOLOGY AND CONSCIOUSNESS SCIENCE

Henry Shevlin, University of Cambridge

Abstract:
Even as the cognitive sophistication of AI systems improves at a rapid pace, debates around AI consciousness remain mired in metaphysical and methodological controversies. From substrate neutrality to multiple realization, the question of whether any artificial system could have subjective experience remains hotly contested, with little common ground between AI consciousness liberals and skeptics. In this talk, I argue that we should not expect these disputes to achieve consensus resolution through new empirical evidence or novel philosophical arguments alone. Instead, I suggest that as the depth and complexity of human-AI interactions increases, intuitive priors among both experts and the general public are likely to increasingly favor theories willing to countenance artificial consciousness. A further question is whether such shifts in sentiment can be rationally justified. I outline the case both for and against the role of intuitions in consciousness science, while emphasising that abandoning intuitions all together would require a radical overhaul of current methods.

Wanja Wiese

6.3 - FROM METAPHYSICAL DISPUTES TO OPEN EMPIRICAL QUESTIONS: AI CONSCIOUSNESS AS A SCIENTIFIC PROBLEM

Wanja Wiese, Ruhr-University Bochum

Abstract:

Can science meaningfully contribute to questions about AI consciousness? If yes, does this presuppose a judgment call about fundamental metaphysical questions?

I suggest treating these questions themselves as empirical scientific questions. To answer them, we first have to identify relevant fundamental issues about the nature of consciousness, such as the following:

i) What features of living organisms, if any, are required for consciousness?

ii) What differences between physical agents (in non-virtual environments) and virtual agents (in computer simulations), if any, matter for consciousness?

I argue that science can meaningfully contribute to these questions, by providing rigorous accounts of potentially relevant features of living organisms, and differences between virtual and non-virtual agents. Specifying these features and differences, ideally mathematically, is the first step towards answering questions like (i) and (ii). Going forward, science can at least contribute to answers by making these questions more intelligible and tractable.

Adrian Wieczorek

7.1 - Why the Frame Problem undermines Situated Cognition and how Situational Framework Analysis might be a new foundation to understand context-sensitivity

Adrian Wieczorek, Technical University of Berlin

Abstract:
In cognitive science and AI research, understanding is strongly related to relevance: intelligent human agents are capable of being drawn to innumerous features of their open-ended worlds but only a small subset of them is relevant for a given, dynamically changing context. However, the problem arises how such dynamic contexts are established in the first place – given the fact that relevance is defined holistically and can come from any direction in the agent's background (Fodor 2000). This "Frame Problem" is usually described in representational terms, often with respect to symbol-based AI and robotics (Dennett 1984; Fodor 2000; Dreyfus 2007). After all, representations of facts are, in themselves, neutral on how they are relevant for a given agent, which makes it hard to derive a frame of meaningfulness from them. Naturalistic or engineering solutions, in terms of "relevance realization" (Vervaeke et al. 2012), are still missing (Froese & Taguchi 2019). Examples from machine learning are generalization error given new, unseen data (Jakubovitz et al. 2019), such as overfitting in LLMs (Mengqi et al. 2024), learning noise and irrelevant correlations. This indicates a lack of understanding in terms of reduction to the relevant data in highly context-variant, tractable and robust ways. To (dis)solve the Frame Problem, proponents of Situated Cognition abandon representations and differentiate between affordances and solicitations: Only the latter call for a certain action and are established through dynamical coupling to specific sensory cues (Dreyfus 2007; Wheeler 2008) and/or relations to skillfulness acquired in shared forms of life and material settings (Rietveld & Kiverstein 2014), often grounded in self-organization principles (Bruineberg/Rietveld 2014; Kiverstein et al. 2022).

This talk has two goals:

First, it is shown that the situated attempts to (dis)solve the frame problem also fail because they all beg the question: Either they still presuppose a principle for individuating situations ("this particular situation", "a context of activity", "behavior setting") the realization of which depends on a solution to the problem. Or the material environments, the mere possession of skills ("abilities") and the sensory modules call again for contexts which govern learning, activate/control the deployment of skills and guide attention to the relevant cues in dynamic ways. Therefore, the Frame Problem is a hard problem still underappreciated by cognitive science. This failure of naturalizing situatedness has serious consequences: It's not just that Situated Cognition is weakened as an alternative to representationalism, if the attunement to relevant affordances remains unclear. Rather, and even more concerning: Without scientific principle of how situations are individuated, and how affordance solicit, two constitutive notions of Situated Cognition itself are undermined.

Second, a new alternative approach is outlined, Situational Framework Analysis (SFA). SFA does not aim to naturalistically ground situations in de-contextualized causal settings (mainly dynamical systems) but rather describes their structural features “from within” in a systematic way. This is not a solution but a first step in terms of understanding how situations are structured and individuated. SFA has three objectives: (1) To shed light on this basic framework in which relevance is realized, four essential structural elements of situatedness (phenomenality, involvement, integration, and appropriation) and their parameters are introduced. (2) In this way, the crucial but rather vague concept of situatedness is provided with a stronger theoretical footing, going beyond mere causal dynamical relations. (3) But this phenomenologically inspired methodological approach remains naturalistically committed: SFA aims to provide the cognitive sciences with a conceptual tool to schematize the contexts of situated tasks and solutions in a more rigorous manner. Notably in psychiatric cases, first-person reports provide evidence that different “fields of affordances” take shape, for example between situations of OCD and depression (Rietveld/Kiverstein 2014; Bruineberg/Rietveld 2014). This is not only in line with rich phenomenological analyses of situatedness in terms of boredom or anxiety (Heidegger 1962). It also serves as an empirical starting point that calls for SFA to analyze systematically the (dis)orders of relevance-sensitivity. This has fruitful implications for various domains of context-sensitive understanding.

David Spurrett

7.2 - On Hostile Epistemic Actions

David Spurrett, University of KwaZulu-Natal

Abstract:
That 'epistemic actions' exist is a key committment of situated cognitive science. On earlier views such as Nilsson's 'sense/plan/act' cycle actions are produced after cognition is completed (Nilsson 1998). Kirsh and others introduced epistemic actions to emphasise that worldly activity can advance cognitive goals, lowering or transforming computational demands by reconfiguring the task environment. In the founding paper on epistemic actions Kirsh and Maglio introduce them as "used to change the world in order to simplify the problem- solving task" (Kirsh and Maglio 1994, 513). Soon after they gloss them as "physical actions that make mental computation easier, faster, or more reliable" and specify that they are actions that "an agent performs to change his or her own computational state" (Kirsh and Maglio 1994, 513-514). In this founding account epistemic actions are by definition (1) self-directed, and (2) beneficial for the agent performing them. Neither of these stipulations is defended by any argument, although subsequent discussion of epistemic actions tends to repeat both restrictions (e.g. Clark 2023). I argue that we should generalise the idea of epistemic actions so that being self-directed, and being beneficial, are possibilities among others. One payoff of generalising the idea in this way is bringing hostile epistemic actions into focus.

First, we shouldn't insist that epistemic actions are always self-directed. A key motivation for the view that cognition is situated is recognition of the cognitive contribution made by things external to the brain, including the activities of other agents. The idea of scaffolding is often credited to Lev Vygotsky, who described a 'zone of proximal development' in which a task could only be performed with external support (Vygotsky 1978). Human infants learning to walk experience a stage where they can take several steps only with some support, which could the hand of a helpful adult or a wall or piece of furniture. The action of offering a hand to a toddler during this stage is a contribution to the toddler's locomotion. More generally, many actions in developmental and instructional settings make sense as other-directed aids to cognition, and are often recognised as cognitive scaffolding. Epistemic actions as traditionally understood can produce scaffolding for their actors — recall Kirsh's cook pre-arranging sliced ingredients to simplify assembling a complex dish (Kirsh 1995). If we accept that the actions of one agent can supportively scaffold the cognition of another, those actions should count as (beneficial) other-directed epistemic actions. They are actions that 'simplify the problem-solving task' or make cognition 'easier, faster or more reliable' for another. When a female firefly produces a signal that helps males of her own species locate her, this is plausibly a benign other directed epistemic action. The males in range of the signal find her more quickly or reliably with her flashing than without it. Refusing to count this as an epistemic action on the basis that epistemic actions must change the agent's "own computational state" separates cases that we should group together on the basis of functional similarity. We should regard an action as epistemic if it tends to change some agent's cognitive state, which doesn't have to be the states of the agent performing the action.

Second, we should recognise epistemic actions that are hostile, not benign. The idea of hostility is from Sterelny, who observes that Brooks's view (e.g. 1991) that the world can be used as 'its own best model' rather than relying on internal representations, "treat[s] the world as benign, or at worst indifferent" (Sterelny 2000, 210). Sterelny argues that "the world is frequently hostile". This is because the world contains other agents, many of them with competing and conflicting interests who stand to gain from deception and manipulation (2000, p. 215). Hostility, then, is the expression of antagonistic or competing interests interests through information. Sterelny (2003) argues that hostility is a key source of environmental complexity, and a significant driver of cognitive evolution.

A hostile epistemic action changes the state of another agent in ways that undermine the target agent's interests, and serves the the agent making the action. The predatory Photuris firefly makes flashes similar to those of the females of other firefly species that males are adapted to approach. The Photuris signals work when they have similar effects on male fireflies of other species, and encourage their approach. The consequences are dissimilar because a same-species female is a mating opportunity, making approach biologically beneficial, whereas approaching a Photuris is courting death. The similarity of mechanism encourages us to group the two cases together as epistemic actions. The difference in details of cause and consequence is what makes the flashes of the Photuris hostile epistemic actions. The same goes for cuckoo chick begging (which works like reed- warbler chick begging), and parasitic beetles releasing pheromones or making food-eliciting gestures (which work like the secretions and movements of ant colony members). The similarity of effect comes with differences in the consequences of being influenced, but that is the point. Sterelny showed us how cue bound behaviour creates an opportunity for exploitation by agents who can use the cues as levers. When this happens it makes cue-guided agents less reliable in tracking the regularities their control systems evolved to track, and more reliable in leading to behaviours that benefit the manipulators. These cases are examples of hostility in Sterelny (2003). My point is that some hostility consists of other- directed epistemic actions. This is not limited to non-human animal cases. In Shakespeare's Othello the villain Iago arranges for a handkerchief that Othello had given to his wife to be seen by Othello in the apartment of Cassio. Iago does this after other manipulations, so that Othello takes it as confirmation that Desdemona and Cassio are lovers. Iago makes multiple hostile epistemic actions when he manipulates the circumstances in which Othello encounters evidence, and what evidences is visible.

When we expand the notion of epistemic actions in this way, where they can be self- or other-directed, and benign or hostile, a fourth possibly surprising option appears, the self-directed hostile epistemic action:

	Self-directed	Other directed
Beneficial	Traditional 'DIY'	Helpful other
Hostile	Self-sabotage	Antagonistic other

The idea of self-directed epistemic hosility raises obvious questions. Why would any agent undermine their own interests? How could any agent be hostile to its own cognition? If we

insist that agents are unified in their interests, such self- undermining seems unlikely or impossible. If we allow that at least some whole agents can be genuinely conflicted, then some interests might act against others, especially if they were dominant at different times. Philosophy has long considered the possibility of self-deception. If self-deception is possible and involves avoiding some evidence and approaching others (e.g. Funkhauser 2005) then it might involve epistemic actions, in which an agent acted against their own epistemic interests.

This argument complements and extends that for 'Hostile Scaffolding' (Timms & Spurrett 2023). It matters to recognise hostile epistemic actions because the interests and goals of many agents can be served by influencing the cognition of others, just as most agents' metabolic goals can be served by ingesting or parasitising other organisms. Sterelny is correct that any model of cognition "that leaves out hostility is not a model of cognition in the wild" (2000, p. 215).

Francesco Marchi

7.3 - THE ABC OF BELIEVING: A THEORY OF FLEXIBLE BELIEF CONSTRUCTION

Francesco Marchi, Ruhr-Universität Bochum

Abstract:
Belief acquisition and update is arguably one of the most important processes of human mental life. Belief attribution is overwhelmingly present both in science and common parlance to understand, explain and predict the behavior of other agents. And yet, we are still far from having a satisfactory understanding of the nature of belief and the dynamics underlying believing. Part of the issue is that, in this debate, there is often talk of beliefs as if they would exist in isolation from other aspects of the mind. It is sometimes neglected that believing is a dynamic situated process that interacts with motivation, emotion, learning and remembering to guide the agent's perception-action routines to successful outcomes. In this paper I propose a novel constructive account of the dynamics of believing that does justice to the most influential views in the field, while overcoming one of the most pressing issues in the debate, namely the problem of belief storage. The problem of belief storage is the problem of specifying how beliefs are stored and selectively re-activated in order to provide the foundation of reasoning and knowledge. If beliefs are taken to be entities that are stored in the mind in some shape or form this could lead to an informational explosion that does not sit well with what we know today about the resource limitations of a biological cognitive system, such as the human brain.

Dominik Battefeld

8.1 - IMPLEMENTING A COMPUTATIONAL COGNITIVE PROCESS MODEL OF MEDICAL DIAGNOSTIC REASONING

Dominik Battefeld, Faculty of Technology, Bielefeld University, Social Cognitive Systems Group, CITEC, Bielefeld University

Stefan Kopp, Faculty of Technology, Bielefeld University, Social Cognitive Systems Group, CITEC, Bielefeld University

Abstract:
Medical diagnoses are an important case of cognitive reasoning processes. While numerous studies have been conducted to understand how a particular physician arrives at a diagnostic decision for a particular patient, researchers like Sox et al. conclude that “we know more about how clinicians should reason than about how they do reason”. We have compiled established theoretical models of diagnostic reasoning and empirical findings on diagnostic reasoning into a preliminary computational cognitive process model that captures how a physician explores information about a patient piece by piece based on their current opinion and active hypotheses to ultimately arrive at a final diagnostic decision.

The core idea behind our model is to frame diagnostic reasoning as a process of self-targeted argu- mentation. Here, argumentation is not understood in the context of a debate but rather as an epistemic practice, where information about the patient (i.e. evidence) is collected to strengthen the justification of one’s own belief. Maximal justification is reached once the strength of the collected evi- dence subjectively warrants the presence of a disease while also subjectively warranting the absence of each considered alternative disease. We implement this process of argumentation optimization in a Markov Decision Process (MDP). Each state captures the currently active hypotheses and the already explored information about the patient. The action space comprises one query action for each unknown information and one commit action for each possible disease to diagnose. The transition model is a deterministic response of the patient to each query and commit actions end the diagnosis. The MDP is solved via an extended version of Monte-Carlo Tree Search (MCTS) to model mentally simulated patient responses with chance nodes in the search tree. In each step of the diagnosis process the evidential strength of new information is quantified by an ex- tended version of the log-likelihood measure, a Bayesian confirmation measure that captures how the belief in hypothesis H changes in the light of evidence et given already explored information (e1, . . . , et−1). The empirical observation that different physicians diagnose the same pa- tient in different ways is thus uplifted to the general phenomenon of differential belief updating, where individuals respond to the same evidence in different ways. This change in belief - summed over all active hypotheses - is used as reward signal within the MDP.

The differences in belief updates between physicians are hypothesized to originate from the difference in subjective probability judgments which are used to calculate the evidential strength. The differences in subjective probability judgments are in turn hypothesized to be a product of differences in the ex- periential knowledge, i.e. which patients a physician has seen and how well they are able to recall characteristics of these patients from long-term memory.

In line with recent criticism on the empirical validity of the classical dual-process theory in diagnostic reasoning , we too argue that the ob- served empirical difference in analytical versus intuitive reasoning styles is less a product of two concurrent reasoning engines System I and System II but rather a function of familiarity. If a pa- tient exhibits a symptom that the physician has seen a thousand times before, the subjective probability judgment will be based on the fast and intuitive exemplar-based knowledge retrieval based on the sim- ilarity between this patient and other previously seen patients stored in memory. If a patient exhibits a symptom that the physician has never seen before, the subjective probability judgment defaults to a slow and analytical abstract knowledge retrieval based on general medical knowledge the physician has learned during medical training. The abstract knowledge retrieval utilizes the Autocorrelated Bayesian Sampler model (ABS) to infer subjective probability judgments based on samples drawn from the MC3 algorithm. The exemplar- based knowledge retrieval is based on the MINERVA model where previously seen patients are stored as memory traces, i.e. vector embeddings that encode the patient's sex, age, diagnosis, risk factors, and symptoms. Subjective probability judgments are calculated through the similarity between embedded observed evidence and all memory traces in parallel which leads to an overall echo intensity that pushes the probability response either towards 0.0 or 1.0 for a negative or positive echo respectively.

Our model combines multiple general psychological theories and empirical evidence on diagnostic rea- soning to form a coherent representation. The predominant hypothetico-deductive reasoning model , where a physician generates hypotheses early on and then strives to verify or falsify them, is incorpo- rated through a rational justification optimization of one's own opinion where the log-likelihood measure has proven to empirically correlate with the perceived evidential strength of lay people. The widely accepted dual-process theory of diagnostic reasoning is rephrased to a dual-route knowl- edge retrieval from memory to match recent criticism and contradictory empirical evidence. Illness-script theory as the basis for fast and intuitive judgments is represented through the exemplar- based memory recall to estimate subjective probabilities and the major emphasis on experiential knowledge in developing medical expertise is captured in the encoding of previously seen patients in the MINERVA model. With each new patient or disease encounter from a specific medical domain, the number of stored memory traces in that domain within MINERVA increases, thereby enhancing the focus and accuracy of the physician's subjective probability judgments. This property is most evident in the fact that MINERVA reduces to Bayes theorem if one assumes each trace to be a perfect copy (i.e. nothing is forgotten) and similarity-based retrieval from memory to only include exact matches. Under these constraints, MINERVA becomes an event counter that produces veridical probabilities. Although theoretical in nature, these considerations connect to empirical observations that experts di- agnose faster, more accurate, and more focused than novices. Even among experts, Walsh et al. report that diagnostic agreement increases with growing experience in the domain of the disease. The current implementation lacks a mechanism to generate or discard hypotheses, which is central to diagnostic reasoning. We plan to integrate this mechanism analogously to the subjective probability judgments, i.e. based on an abstract or exemplar-based knowledge retrieval depending on the familiar- ity with a particular disease. We additionally plan to empirically validate the claims made by the model first by aligning the predicted subjective

probability judgments to recorded estimates of physicians and by comparing the generated reasoning trajectories (i.e. sequential information queries and hypothesis updates) to monitored diagnosis processes.

Amelie Sophie Robrecht

8.2 - A Computational Approach to Adaptive Explanation Generation Based on Cognitive Partner Models

Amelie Sophie Robrecht, CITEC, Bielefeld University, TRR 318 - Constructing Explainability

Stefan Kopp, CITEC, Bielefeld University, TRR 318 - Constructing Explainability

Abstract:
When explaining, people often adjust their explanations by modifying their vocabulary, pace, or the way they present information to align with their interlocutor. This ability to adapt can even be observed when interacting with strangers, indicating that the process is quick and instinctive. Furthermore, adap- tation can occur without directly communicating partner-specific information, such as the listener's level of understanding (LoU), attentiveness, or domain-expertise. This shows that humans can gauge their interlocutor's needs from implicit cues and the corresponding adaptations enhance explanatory success and efficiency. For example, an expert might use specialized comparisons to shorten their explanations when speaking to another expert, whereas it could confuse someone who is unfamiliar with the topic. In our research, we develop an artificial explainer (SNAPE-PM) that models the current user to efficiently tailor its explanations. The agent's architecture is inspired by patterns observed in human- human explanations. We define an explanation as an interactive process between an explainer (ER) and an explainee (EE), where both collaborate to ground an explanandum (EM) in a co-constructive manner. The board game Quarto serves as the EM, and the explanations are generated in German. Adapting an explanations requires an understanding of the relevant states of the interaction partner, i.e., building a suitable partner model (PM). Although it is well known that various features are essential for a suitable PM, current computational approaches primarily focus on modeling the user's knowledge. However, we argue that additional features are also important to consider for adaptive explaining, where the interaction goal is to ground the EM. We consider four features as particularly relevant: exper- tise, cognitive load, attentiveness, and cooperativeness. For example, a comparison such as the game is just like Tic Tac Toe can significantly shorten an explanation if the user possesses the necessary level of expertise to understand the analogy. Conversely, if the user is not familiar with the game, such a com- parison can lead to confusion, negatively affecting the explanation efficiency. Another crucial feature for adapting information complexity is cognitive load. If the user is experiencing high cognitive load, infor- mation should be introduced one piece at a time; in contrast, a user with low cognitive load benefits from receiving multiple information in one statement, preventing boredom and mind-wandering. Furthermore, a user displaying high cooperativeness is likely to inform the ER about misunderstandings. While, if cooperativeness is low, the ER must gain information about the LoU using verification questions. These examples illustrate that the four PM features are vital when planning explanations, and we conjecture that different value combinations of these features lead to different explanation strategies being optimally adapted to the respective EE's. Based on this assumption, the rational explainer SNAPE-PM is designed to build and dynamically update a PM in order to adapt its explanation strategy in real-time without requiring any previous infor- mation. An explanation consists of multiple turns that comprise potential user feedback and the agent's next utterance presented as text

on a graphical user interface. Users can provide either positive or negative nonverbal feedback by clicking on a nodding or frowning smiley, or they can type in a question in a textbox. The latter is processed by a fine-tuned version of Gemma3:27b, optimized to translate the natural language question into a question type (polar or open) and identify the requested information (as triple in a knowledge graph). This triple-move combination is sent to a Model Update component, which disseminates it to the core SNAPE-PM components: The Graph Data Base, implemented in Neo4J, updates the LoU of the current triple based on the received feedback. This labeled knowledge graph acts as a memory component for the user's knowledge. Additionally, a Bayesian Belief Update updates the PM from the feedback received. It consists of a dynamic Bayesian network (DBN) with four hidden and four observable features. The hidden features can be ei- ther static (like expertise) or dynamic (such as cognitive load, attentiveness, and cooperativeness), with the estimation of all features being inherently dynamic. The user's expertise indicates their general knowl- edge of the domain—in this case, board games—and increases with the amount of positive feedback received. Connected to expertise is cognitive load, which is observed through variations in the user's typing speed. The features attentiveness and cooperativeness relate to the different functions of linguis- tic feedback. Cooperativeness captures higher functions like understanding and attitude, increasing only with substantive feedback, while attentiveness rises with any feedback, addressing functions of contact and perception. Based on the DBN-based state estimates, features values are passed to a Decision Making Component, which determines the next action (what to say) and the manner of delivery (how to say it). This component is formalized as a non-stationary Markov Decision Process (MDP). To maintain real-time capabilities, the system breaks down the domain into semantically related expla- nation blocks, as observed in human-to-human interactions [5]. Depending on the current context and PM, the system can decide between four different actions (1) introduce a new information, (2) deepen previously introduced but ungrounded information, (3) answer a question, or (4) transition to the next explanation block. The move selection (how to verbalize an action) relies on estimated PM features, which are factored into the calculation of MDP transition probabilities, expected increases in the level of understanding, and rewards. The MDP is solved using Monte Carlo Tree Search (MCTS), and the selected triple and move are sent to another fine-tuned Gemma3:27b optimized for natural language generation. The resulting utterance is then presented to the user. The modular design of SNAPE-PM and its core components (DBN, knowledge graph, MDP) make the process transparent and increases explainability. At the same time, it allows for the replacement of single components to accommodate different explanation needs, including swapping the graph database, utilizing different LLMs for natural language understanding or generation, or substituting the DBN with alternative PM features. The code for SNAPE-PM is available on GitHub: https://github.com/arobrecht/severus-study. To further increase the explainability of SNAPE-PM, we added a visualization component that allows for tracking decisions and the reasons behind those in real-time. The visualization consists of three parts: (1) a knowledge graph representing the current explanation block, where colors indicate the estimated LoU; (2) the PM feature varying over time; and (3) the top five actions that received the highest reward when solving the MPD. We are currently conducting an online ablation study to examine the effects of PM features compared to a PM that solely tracks estimated knowledge. Additionally, we are testing both conditions against a non-adaptive baseline. A previous study on a former version of the agent indicated that deep enabledness – the ability

to transfer and apply knowledge – is enhanced when generating user-adaptive explanations, while satisfaction is not. However, findings regarding the effects of adaptive interactions are mixed: while some studies report negative effects, others show a positive impact on the performance. Most of those studies report a positive impact on the user's satisfaction. Therefore, our study will focus on assessing the impact on (1) objective understanding, (2) the perception of the explanation, and (3) the perception of the ER.

Norbert Fürstenau

8.3 - Evidence for Hill-function power law of subjective mental workload using air-traffic control human-in-the-loop simulation

Norbert Fürstenau, German Aerospace Center, Inst. of Flight Guidance

Abstract:
In recent years, new technological and operational developments have led to the introduction of Remote Tower Operations (RTO) into air traffic control. In RTO, air traffic control officers (ATCOs) work from newly designed work environments independent from conventional tower buildings and can control several airports from afar [1] [2]. This change in the ATCOs' workplace and the necessary new human-system interfaces can lead to changes in human performance which need to be assessed and evaluated, preferably in a non-intrusive and objective way. Human performance and cognitive demand can be quantified through objective physiological measures such as heart rate and variability, electroencephalography, functional near-infrared spectroscopy [3] [4], and subjective methods like the "instantaneous self-assessment" (ISA) [5] [6] [7].
In the present work, I argue that such subjective ratings can be used to infer objective task demands as well. I provide new evidence for a Hill-function type of power law relationship between objective cognitive task load (TL) and subjective mental workload (MWL) variables by means of a human-in-the-loop (HITL) air traffic control simulation experiment [8] [9]. Hill derived his empirical hemoglobin oxygen (HbO_2) equilibrium concentration function [10] [11] ($R = R_{max} / (1 + 1 / (K L^m))$, $L = O_2$–ligand concentration, K = reaction constant, m = effective number of saturated O_2-attachment sites) to simplify parameter estimates of experimental saturation data [12] [13]. In the following, I show that results of our MWL(TL)-data analysis agrees with Hill's HbO_2-dissociation curve fit in [10]. In our experiment 12 professional ATCOs controlled simulated traffic in an RTC-operators dual-airport work environment (Fig.1; [1] [2]).

Figure 1: Controllers workplace in Remote Tower Center simulation environment with standard equipment (flight strips, weather display, control zone radar) and video panorama system that replaces out-of-windows view for two remote airports. Microphone for online radio communication with (pseudo) pilots in a separate room driving the simulated air traffic.

In the present work I analyze data only from single ATCOs controlling a single airport. Each ATCO completed eight simulation runs with a duration of 25 min each. Within runs traffic level n varied as independent load variable. Subjective MWL(n) was assessed every 2 min by means of the ISA (one-dimensional 5-point Likert scale, from underload $I_d = 1$ to overload $I_u = 5$). Objective TL(n) data were derived from ATCO's communication with pilots (frequency of radio calls, RC) [8] [9] [14]).
Based on the assumption of limited cognitive and energetic resources [15] [16], and with reference to an empirical approach to formalize HITL-data by a sigmoid characteristic [6] [17],

a formal logistic resource limitation model (LRL) had been derived for estimating model parameters of measured ISA-MWL(n) and RC-TL(n) characteristics [9] [18]. Prior knowledge on ISA and RC scaling was used (lower and upper sigmoid (asymptote) levels $I_d = 1$, $I_u = 5$, MWL-bias D := I_d, R(n)-sigmoid inversion at R(n=0) = $R_0 = 0$, i.e. nearly linear increase of R(n) ≥ R_0 for vanishing traffic) to minimize number of LRL-model parameters for theoretical prediction and

$$I(n) = \frac{I_u - \Delta}{1 + exp\left\{-\frac{n-\mu}{\nu}\right\}} + \Delta \qquad (1)$$

estimation by nonlinear regression:

$$R(n) = R_u \left(\frac{2}{1 + \exp\{-\frac{n}{\rho}\}} - 1 \right) = R_u tanh\left(\frac{n}{2\rho}\right) \qquad (2)$$

Sensitivity parameters n and r determine the slope of the sigmoid curves whereas m quantifies the traffic load at the inversion point. Logistic shift factor k can be expressed by intersection I(n=0) := I_0 and MWL-bias D as k = exp(m/n) = $(I_u - I_0)/(I_0 - D)$, and may be a large number due to $D - I_0 << 1$. Figure *2* shows the experimental results of the simultaneous subjective ISA-MWL and objective RC-TL data (averages across participants and repeated simulation runs) as dependent on traffic count n, together with nonlinear 2-parameter (I_0, m) and (R_u, r) lsq.-regressions.

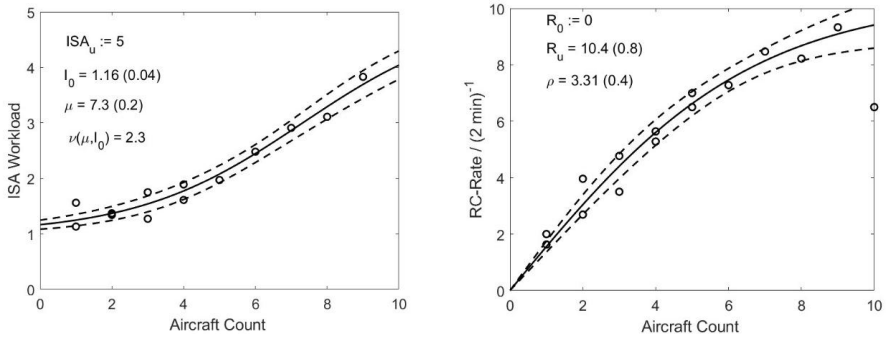

Figure 2: MWL-reporting and simultaneous RC-frequency during RTC-HITL simulation (2 min intervals; circles: averaged data), as dependent on traffic count n (aircraft within sector). Solid lines: Robust logistic 2-parameter lsq.-fits with Matlab®-Nlinfit (bi-square residuals weighting for outlier suppression, 95% CI (dashed)). Left: subjective ISA-MWL(n). Parameter estimates (± SE): I_0, m, with n(I_0, m) calculated. Right: RC-TL(n) with estimates R_u, r (± SE).

Combination of logistic functions (1) and (2) through replacement of the traffic load variable ($n = r \ln[(1 + p_R)/(1 - p_R)]$, $p_R = R(n)/R_u$)) in equ.(1) leads directly to a two-parametric ISA(RC)-power law with exponent g = r/n as ratio of logistic sensitivity parameters. With normalized ISA-MWL variable $I(R)/I_u := p_I(R)$ the power law is given by:

$$p_I = \frac{1}{1 + kS^{-\gamma}} + \frac{\Delta/I_u}{1 + \frac{1}{k}S^{\gamma}} = \frac{1 + \delta K_c^{-1} S^{-\gamma}}{1 + K_c^{-1} S^{-\gamma}} \qquad (3)$$

with $0 \le p_I \le 1$, $d = D / I_u = 0.2$. Power law model equ.(3) exhibits formal correspondence to Hill's empirical function (see above), with small additional second term < d. Power law parameters predicted from LRL-parameters (g = 1.2 (±0.04) and $K_c = 1/k = 0.042$ (±0.01)) may be tested by comparison with Nlinfit-estimates using cognitive Hill-function (3). The new cognitive TL variable S with $1 \ge 1/S \ge 0$ has the form of a contrast function of the radio calls variable $(R_u - R)/(R_u + R) = (1 - p_R)/(1 + p_R) := 1/S(p_R)$, with $0 \le p_R \le 1$. 1/S depends on the remaining capacity for radio calls (R_u - R). Interestingly, the inverse shift factor $1/k = (I_0 - D)/(I_u$

– I_0) := K_c corresponds to Hill's HbO_2 equilibrium reaction constant K while exponent g ~ m. For obtaining regression estimates of power law parameters g, k we use LRL-model prior information, with R_u := 10.
As expected, parameter estimates (1.2 ± 0.1, $K_c = (I_0 - D)/(I_u - I_0)$ = 0.04 ± 0.01) agree with those obtained with the LRL model within uncertainties, and may be compared with Hill's classical results. Hill showed that his empirical function for fitting experimental HbO_2-saturation data provided a good approximation, also compared with a more realistic physicochemical 8 parameter model derived from the mass-action law ([10] [12] [13], equilibrium reaction constants K_r and weighting factors a_r, r = 1,…,4 = Hb-attachment sites for O_2). In [10] Hill reported parameter estimates m = 1.4, K = 0.015. Within uncertainties these results are in surprising agreement with our cognitive power law parameters (g, K_c), considering the completely different origin of our psychological/cognitive and Hill's physicochemical datasets. Using the present parameter set we derived in [19] an explicit Hill-type equilibrium relationship between subjective ISA-MWL and oxyhemoglobin concentration to provide a possibility for testing our MWL-model with experimental oxygenation data.

It appears worth mentioning that the exponent g also has the typical order of magnitude of Stevens power law of psychophysics $P(S) = K\ S^g$ which he had validated through sensation matching experiments for various physical stimuli of rel. intensity S ([20] [21], e.g. acoustic loudness). Formal correspondence to the cognitive power law $P = K_c\ S^g$ was obtained by transforming the MWL-ISA variable into $P_I(S) = (I - D)/(I_u - I) \geq 0.053$ for I_0 = 1.2, with RC-TL variable $S = (1 + p_R)/(1 - p_R) \geq 1$ [18] [9].

I am indebted to Anne Papenfuss for providing the preprocessed data and Anneke Hamann for reviewing and improving the first manuscript version.

REFERENCES

[1] N. Fürstenau, Virtual and Remote Control Tower, N. Fürstenau, Ed., Springer, 2016.

[2] N. Fürstenau, A. Papenfuss and J. Jakobi, Eds., Virtual and Remote Control Tower, 2nd edition, Switzerland: Springer Nature, 2022.

[3] S. Loft, P. Sanderson, A. Neal and M. Mooij, "Modeling and predicting mental workload in en route air traffic control: Critical rewiew and broader implication," *Human Factors,* vol. 49, pp. 376-399, 2007.

[4] A. Hamann and N. Carstengerdes, "Investigating mental workload induced changes in cortical oxygenation and frontal theta activity during simulated flights," *Scientific Reports,* vol. 12, no. 6449, 2022.

[5] C. S. Jordan and S. D. Brennen, "Instantanous self-assessment of workload technique (ISA)," Portsmouth, GB, 1992.

[6] P. U. Lee, "A non-linear relationship between controller workload and traffic count," *Proceedings Human Factors and Ergonomics Society,* vol. 49, pp. 1129-1133, 2005.

[7] E. Stein, "Air traffic controller workload: An examination of workload probe," DOT/FAA/CT-TN84/24, Atlantic City, NJ, 1985.

[8] M. Lange, C. Möhlenbrink and A. Papenfuß, "Analyse des Zusammenhangs zwischen dem Workload von Towerlotsen und objektiven Arbeitsparametern," German Aerospace Center (DLR), Internal Report IB112-2011/46, Braunschweig, 2011.

[9] N. Fürstenau and A. Papenfuß, "Power law model of subjective mental load and validation by Remote Tower Center simulation," in *Virtual and Remote Control Tower*, 2 ed., N. Fürstenau, A. Papenfuss and J. Jakobi, Eds., Cham, Switzerland, Springer Nature Publishers, 2022, pp. 293-342.

[10] A. V. Hill, "The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves," *Proceedings Physiol. Soc.,* vol. 40, pp. IV - VII, 1910.

[11] R. Gesztelyi, J. Zsuga, A. Kemeny-Beke, B. Varga, B. Juhasz and A. Tosaki, "The Hill equation and the origin of quantitative pharmacology," *Arch. Hist Exact Sci.,* vol. 66, pp. 427-438, 2012.

[12] J. Barcroft and M. Camis, "The dissociation curve of blood," *J. Physiology,* vol. 39, no. (5), pp. 118-142, 1909.

[13] J. Barcroft and A. Hill, "The nature of oxyhemoglobin, with a note on its molecular weight," vol. 39, no. 6, pp. 411-428, 8 March 1910.

[14] J. Djokic, B. Lorenz and H. Fricke, "Air Traffic Control Complexity as Workload Driver," *Transportation Research Part C,* vol. 18, pp. 930-936, 2010.

[15] W. Laskowski und W. Pohlit, Biophysik, Stuttgart: Georg Thieme Verlag, 1974.

[16] N. Birbaumer and R. F. Schmidt, Biologische Psychologie, 7. ed., Heidelberg: Springer Medizin Verlag, 2010, pp. 214-221.

[17] P. U. Lee, J. Mercer, N. Smith and E. Palmer, "A non-linear relationship between controller workload, task load, and traffic density: The straw that broke the camel's back," *Proceedings Int. Symp. Aviation Psychology,* pp. 438-444, 2005.

[18] N. Fürstenau and T. Radüntz, "Power law model for Subjective Mental Workload and validation through air-traffic control human-in-the-loop simulation," *Cognition, Technology, and Work,* vol. 24, no. 2, pp. 291-315, 2022.

[19] N. Fürstenau, "Hill function of oxyhemoglobin saturation quantifies power law parameters of subjective mental workload," *Cognition, Technology, and work, under review,* 2025.

[20] S. S. Stevens, Psychophysics: Introduction to its perceptual, neural and social prospects, New York: Wiley, 1975.

[21] S. W. Link, The Wave Theory of Difference and Similarity, New York: Lawrence Erlbaum Associates and Routledge, 1992.

Witold Kraszewski & Katarzyna Skowrońska

9.1 - ROLLING WITH THE RHYTHM: INVESTIGATING GROUP BEHAVIOR AND HEART RATE SYNCHRONY IN CLASSROOMS

Witold Kraszewski, University of Warsaw

Katarzyna Skowrońska, University of Warsaw

Julia Słowicka, University of Warsaw

Agnieszka Karpińska, University of Warsaw

Abstract:
Our study stems from a need to understand the dynamic and embodied nature of learning processes in educational settings. Traditional approaches to studying classroom behavior often focus on static outcomes—such as test scores or time-on-task—while overlooking the subtle, moment-to-moment interactions and bodily cues that shape how learning unfolds in real time. Embodied research brings attention to the physical, emotional, and physiological dimensions of classroom life, viewing learning not just as a cognitive activity, but as a deeply social and bodily experience.

While studies of embodied cognition often focus on components of cognition like memory or attention (Ianì, 2019), we take a closer look at classroom dynamics. This is a very convenient format to study, because of its repetitive structure, a 45 minute format and a consistent group of people who gather everyday to perform these lessons. We wanted a way to capture interactions between students as well as student - teacher interactions. To do so we used a multimodal approach using video recordings, heart rate measurements and post lessons questionnaires.

Furthermore, we were interested in the quality of the lesson and its links to participants' emotions. Some studies were done on emotions and stressors in the classroom (Ahmed et al., 2010; Donker et al. 2020). We build on that and provide some objective measurements using HR monitoring and compare it to the questionnaire data we've gathered.

Additionally, we measure heart rate coordination between the participants using Cross Recurrence Analysis methods. Physiological synchrony was stipulated to be a marker for common attention (Dikker et al., 2017) and teacher - student closeness (Bevilacqua et al., 2019) in EEG studies. Smartbands are a cheap, easy to mount and almost out-of-the box way to gather heart rate data, which makes it very much applicable in education studies. We will present some insights into pros and cons of using this solution and how to tackle some of the inherent problems with these not very precise instruments.

Julia Wolf

9.2 - INTRACTABILITY IN SOCIAL COGNITION – THE ROLE OF MINDSHAPING FOR THE ONTOGENETIC DEVELOPMENT OF MINDREADING

Julia Wolf, Ruhr-University Bochum

Abstract:
One of the important questions for an account of mindshaping is how mindshaping relates to mindreading. Both theories of mindreading and mindshaping agree that folk psychology (or the attribution of mental states to oneself and others) plays an important role in social cognition. However, they disagree about the function of folk psychology. Advocates of mindreading assume that folk psychology is a descriptive practice through which we attempt to determine what mental states the other person has in order to predict and explain their behaviour (e.g. Carruthers, 2020; Premack & Woodruff, 1978). Advocates of mindshaping, in contrast, emphasise the normative or regulative dimension of folk psychology, arguing that we use folk psychology to shape the minds and behaviours of others (McGeer, 2007, 2015; Zawidzki, 2008, 2013, 2018). While in the literature, mindshaping is often presented as an alternative to mindreading, a further key argument is that mindshaping is required in for mindreading to be possible. In this paper I will critically evaluate this claim, focusing on ontogenetic development. I begin by considering one of the main arguments for the primacy of mindshaping, namely the intractability of mindreading. I then consider two ways in which an advocate of the mindreading view can respond to this argument, using this as a means to further clarify the distinctive contribution of mindshaping to the development of social cognition. Ultimately, I conclude that the main benefit of mindshaping is that it distributes the cognitive load over a group, rather than only reducing complexity at the level of the individual. Finally, I end with an outlook on what this means for the interaction between mindreading and mindshaping beyond the early years of ontogenetic development. One of the main arguments for the primacy of mindshaping is the Intractability Argument: in the absence of mindshaping, descriptively attributing mental states in order to explain and predict the behaviour of others poses an intractable challenge to the mindreader (Zawidski, 2013). Due to the holism of the mental, any behaviour could have been caused by a potentially infinite number of mental states, while a particular mental state could give rise to a large number of different behaviours (Andrews, 2017; Zawidzki, 2013). Individual variation further complicates matters. Mindshaping, Zawdiski argues, can resolve this intractability as it is able to narrow down and constrain human minds and behaviours. I argue that while this argument has primarily been developed with regards to phylogenetic development, it is if anything stronger when applied to the case of ontogenetic development. I consider two ways in which the defender of mindreading could respond to this. Firstly, they could question whether mindreading really is so intractable. Zawidzki's (2013) argument that rests on the idea that mindreading needs to be accurate and that it would be difficult, if not impossible, to accurately determine the mental states of others. However, it might be disputed that mindreading really needs to be accurate in a way which makes mindreading intractable. Westra (2020), for example, has pointed out that assessing the accuracy of mindreading is complex and that we currently do not have a good way of determining whether mindreading is accurate or not. Against this it might be objected that while we do not know exactly whether mindreading is

accurate and to what degree, we know that it at least needs to have a level of accuracy that is sufficient in order to allow for cooperation in society. The question then is whether this would require a level of mindreading which is intractable, or whether this would be possible in terms of very broad and general mental state attributions. This would mean that mindreading need not always be complex and intractable and therefore might not require prior mindshaping. This, however, is not to deny that there are highly complex forms of mindreading, but that there may be some forms of mindreading which are less complex and intractable. The second objection which a critic of the priority of mindshaping might pose is that there are other ways of reducing the demandingness of mindreading. For example, Peters (2019) argues that biases and stereotypes can play the role of narrowing down the range of possible mental states to be ascribed. Similarly, it has been argued that background information can help to make mindreading easier (Wolf et al., 2023). There are ways, however, in which the advocate of mindshaping can respond to this. Firstly, it can be argued that stereotyping and use of background information just are further instances of mindshaping at play. For example, that people conform to stereotypes because they have been shaped to do so, thus making their prediction in terms of stereotypes so successful. Secondly, it could be argued that stereotypes and background information reduce the complexity of mindreading, but do so at the cost of bringing in complexity in other places. Take the example of using background information in order to disambiguate a scene for mindreading. In a sense this makes the mindreading less computationally demanding on the child as the scope of potential mental states is reduced. However, this reduction in complexity comes at the cost of the child having to be able to remember a wide range of background information and apply this correctly to limit the possible mental states. One way of putting it is that on mindreading views the cognitive load is at the level of the individual. Mindshaping, however, distributes the challenge of mindreading (and the cognitive load it brings with it) across the group – the reduction of possibilities does not happen at the level of the child who can rule out alternatives by using further information, but at the level of the group as a whole through an increased uniformity in behaviour. This is important especially from the perspective of a child learning social cognition: rather than the child having to do all the work to learn the rules of social cognition, much of the work is already done by a community of mindshapers.

Christian Kliesch

9.3 - A RADICALLY EMBODIED TRAJECTORY TOWARDS HUMAN INTERSUBJECTIVITY

Christian Kliesch, Department of Psychology, University of Potsdam

Abstract:
The human ability to interact, feel, and coordinate with others has frequently been considered an important aspect of a uniquely human phenotype. This ability to understand and extend ourselves into others comes with specific human phenomelogical qualities and provides the foundation for other important social skills, such as language and culture. Radically embodied cognitive (REC) accounts of cognition frequently argue that human social cognition is based on notions of intersubjectivity, rather than the explicit representation of mental states. REC proponents argue that others' behaviours are fundamentally accessible to an observer. However, the notion of a uniquely human intersubjectivity has been notoriously difficult to define from an REC perspective. Whereas some descriptions of intersubjectivity are too general to explain its uniquely human aspects, others appeal to vague cognitive differences to explain its emergence. I suggest that these problems can be solved if we take into account humans' unique ontogenetic trajectory. In line with general principles of an enactive and ecological perspective of the organism as a self-organising system that operates within the constraints posed by its ability to act on its environment (Varela et al., 2017), I propose that human ontogeny fundamentally restructures how infants are able to interact with the world, prioritising action and exploration through others over acting and exploring the world on their own (Kliesch, 2025).

Martina Penke

10.1 - VISUAL ATTENTION INFLUENCES CHILDREN'S ORDER OF MENTION IN CONJOINED NOUN PHRASES BUT NOT IN TRANSITIVE SENTENCES

Martina Penke, University of Cologne

Sarah Dolscheid, University of Cologne

Abstract:

The way speakers produce an utterance is closely linked to their allocation of attention. For instance, when asked to describe a picture with two characters, speakers are more prone to start an utterance with a visually cued character which is in the spotlight of their attention (e.g., Gleitman et al., 2007). However, although links between attentional orienting and language production have been attested for adult speakers, this relationship has not yet been investigated in children. In the present study, we shed light on this open issue by testing 4- to 5- year-old children in two language production tasks: a noun pair naming task and a picture-based event description task. In both tasks, children's gaze patterns were monitored via eye-tracking. Methods: In total, 35 German-speaking children (16 female, 19 male) were tested. In the noun pair naming task, children were asked to describe pictures of two characters presented next to one another by producing a conjoined noun phrase (e.g., a dwarf and a clown). In the event description task, children were asked to describe pictures of transitive events in which an agent acted on a patient (e.g., a clown pushing a dwarf). We manipulated children's allocation of attention by means of a brief visual cue (a small red dot) presented for 700 milliseconds in the place where the left character/the patient was about to appear (see Fig. 1). In a baseline condition, no cue was presented. Results: In both tasks, visual cueing was highly effective in modulating children's attention, as revealed by a significant increase in first fixations to the cued character compared to baseline (noun pair naming task; z.ratio = 6.85, $p < .0001$, event description task: z.ratio = 7.34, $p < .0001$). In the noun pair naming task, children were also more likely to start their utterances with the cued character compared to baseline (z.ratio = 2.87, $p = .004$), however, the same did not apply to the event description task (z.ratio = 1.49, $p = .13$, ns). Discussion: Our findings show that the effect of attentional cueing differs depending on the syntactic structure children produce. While attentional orienting affects children's order of mention during the production of simple conjoined noun phrases, the same does not apply to the production of event descriptions that require the production of more complex syntactic structures. Specifically, starting the sentence with the cued patient character would have required to produce a passive sentence or a sentence with topicalized patient (OVS), the acquisition of which we assessed by an act-out-task. Taken together, the data show that the propensity to verbalize what is in the spotlight of attention is modified by syntactic complexity, hence demonstrating a complex interrelationship between attention and language production in children.

Moritz Bammel

10.2 - Exploring the Potential of Recurrence-Based Eye Movement Measures for Mobile and Webcam-Based Applications in Reading Research: Effects of Sampling Rate

Moritz Bammel, Leuphana Universität Lüneburg

Monika Tschense, Leuphana Universität Lüneburg

Sebastian Wallot, Leuphana Universität Lüneburg

Abstract:
Recent work on nonlinear eye movement dynamics during reading has demonstrated the potential of using recurrence-based eye movement measures to predict reading comprehension: Whereas traditional fixation-based measures did not predict reading comprehension reliably, it has been shown that more flexibility in nonlinear eye movement dynamics is associated with more proficient reading comprehension (Bammel & Sanches de Oliveira, 2023; Tschense & Wallot, 2025). However, these findings have been obtained from laboratory experiments using stationary eye-tracking with high temporal resolution (1000 Hz or 500 Hz). Here, we want to explore the feasibility of utilizing recurrence-based eye movement measures in more applied settings employing mobile or webcam-based eye-tracking with lower sampling rates. Previous comparisons of stationary versus mobile and webcam-based eye- tracking systems have shown that the data quality of the latter is comparable to stationary systems in many eye-tracking tasks (Dowiasch et al., 2020; Kaduk et al., 2024), but it remains unclear if this also applies to nonlinear eye movement dynamics that are typically assumed to be found at fast timescales, especially in the context of reading research (Trasmundi et al., 2022; Tschense & Wallot, 2022b). Hence, we aim to investigate whether nonlinear eye movement dynamics at slower timescales – those timescales that can, in theory, be captured by mobile or webcam-based eye-tracking – are indicative of reading comprehension proficiency as well.

Erik Ayari & Manuel Traub

10.3 - From Egocentric Snapshots to Holistic Object Codes: A Contrastive Account of Infant Learning

Ayari Erik, Cognitive Modeling, University of Tübingen

Traub Manuel, Butz Martin V

Abstract: Infants form stable object concepts from egocentric snapshots that – on their own – only provide partial information about the three-dimensional structure. They do so by manual exploration, effectively generating their own learning curriculum. In this work, we model the formation of holistic object representations from similar learning curricula by means of a self-supervised deep learning architecture. Our model yields highly structured representations with emergent properties, abstracting away from single object instances to object categories. This is demonstrated in competitive object classification performance, despite the fact that our model was never presented with any categorical information during training.

Lutz Wehrland

11.1 - RHYTHMS BEYOND WORDS: USING DEEP LEARNING TO EXPLORE TEMPORAL STRUCTURES IN JACKDAW VOCAL COMMUNICATION

Lutz Wehrland, Neural Basis of Learning, Faculty of Psychology, Ruhr University Bochum

Jonas Rose, Neural Basis of Learning, Faculty of Psychology, Ruhr University Bochum

Abstract:
Vocal communication is one of the most remarkable abilities humans have evolved. Char- les Darwin saw in birdsongs the seeds of human language: "The sounds uttered by birds offer in several respects the nearest analogy to language." (Darwin, 1871). Over 150 years later, while human communication is extensively studied, the evolutionary origins of our lan- guage remain unclear (Knight et al., 2000; Hauser et al., 2014; Christiansen & Kirby, 2003). Among primates, we humans occupy a unique position with our sophisticated vocal commu- nication systems (Hauser et al., 2002).

Corvids, renowned for their advanced cognition (Emery, 2004; Emery & Clayton, 2005; Tay- lor, 2014) and belonging to the songbird suborder (Passeri), are rarely recognised for their vocal complexity and thus often overlooked in vocal studies. Yet, they follow temporal rules in vocal exchanges (Kondo et al., 2010), exhibit volitional vocal control (Brecht et al., 2019), and can recognise individuals by call (Benti et al., 2019; Martin et al., 2022). Their vocal rep- ertoires remain underexplored and often subject to human bias (Kondo et al., 2010; Martin et al., 2024). Unlike typical songbirds, corvids do not clearly engage in song, duetting, or cho- rusing, but display a rich variety of calls and mimicry used in complex social contexts (Eng- gist-Dueblin & Pfister, 2002; Kondo & Hiraiwa-Hasegawa, 2015). This makes them a com- pelling model for studying the evolutionary roots of complex communication.

In this talk, I will present findings from my PhD research, where I used deep learning to analyse the temporal structure of jackdaw vocal communication and behaviour, and to build the fundamentals for artificial vocal interaction with our birds. In these studies, we used deep learning convolutional neural networks to detect and track the vocalisations of jackdaws in real time over the course of a full day, and were able to show how the vocal behaviour of jackdaws changes over the day and which factors might influence it, based on statistical analysis using generalized linear mixed models (GLMMs). Additionally, I will talk about how our AI system can play back sounds in response to jackdaw vocalisations, mimicking “real” jackdaw behaviour. Using GLMMs, we analysed the playback experiments and found that our birds vocalised more when a jackdaw call was played back, and less when a neutral sound like wind was presented. Furthermore, the models showed that jackdaws responded faster and more precisely to jackdaw call playbacks compared to neutral stimuli. These results not only advance our understanding of temporal and general corvid vocal be- haviour, but also demonstrate the potential of deep learning systems to facilitate real-time, interactive, and ecologically valid communication experiments. By moving beyond static playback and towards responsive systems, we open new ways to study animal communica- tion in more naturalistic and socially relevant contexts.

Oryan Zacks

11.2 - THE EVOLUTION OF IMAGINATION: A NEURAL TRANSITION FROM REPTILES TO MAMMALS

Oryan Zacks, Tel Aviv University

Abstract:
Human imagination is a complex, rich and highly developed cognitive ability, constituting an important part of human experience (Abraham, 2020). It is a critical component of human linguistic abilities and was probably an evolutionary pre-requisite to developing symbolic language (Dor, 2015). Furthermore, a well-functioning imagination is critical for mental health and well-being, and foundational for creativity and empathy. Nevertheless, imagination is critically understudied as such, with little systematic, in-depth, or interdisciplinary research addressing it directly. Approaching human imagination from an evolutionary perspective, this research will address the following questions: how could such an ability evolve? What were the circumstances that led to the elaboration of this ability and what precursors needed to be in place? Finally, is imagination uniquely human, or is this ability more widespread across the animal kingdom? And if it is more widespread, how can we study non-human imagination?

According to Tomasello (2014, p. 9), "imagining is nothing more or less than the "off-line" simulation of potential perceptual experiences." Such a definition can theoretically be applied to non-human animals, although the major challenge is how to measure or test such mental simulations in the absence of personal verification using human language. Moreover, it is clear that a complex nervous system must be in place for an animal to simulate potential perceptual experiences "off-line" and leverage such simulations in an advantageous manor. Interrogating the neural structures that enable imagination is a critical step in understanding the nature of imagination itself and to uncover possible treatments when imagination is impaired.

A core neural network supporting imagination has been identified in the human brain (Schacter et al., 2012). This network consists of the hippocampus and several neocortical areas found in the medial temporal and frontal lobes, posterior cingulate and retrosplenial cortex, and lateral parietal and temporal areas. This paper will trace the development of these structures throughout vertebrate evolution, identifying key transitions that could determine the presence and extent of imaginative abilities. A critical point in the evolution of this core network is the transition from reptilian-like ancestors to mammals. Not surprisingly, this strongly coincides with a cognitive and behavioural shift. Behaviours that are associated with imagination, such as planning, theory of mind, inference, and episodic-like memory have been found in a diverse range of mammals but have not been reported in reptiles (Zacks et al., 2022).

The neocortex is widely considered the defining feature of the mammalian brain, as its multi-layered structure represents a significant shift from the three-layered reptilian cortex (Shepherd, 2011). The unique neocortical organization imparts computational abilities that are pivotal for the development and maintenance of imagination. The mammalian hippocampus is widely considered as an "ancient" or "conserved" structure, since it maintains a three- layered architecture. However, a closer look suggests that the mammalian

hippocampus has undergone changes on par with those of the neocortex. Furthermore, it occupies a key position within the processing hierarchy of the mammalian brain and is critical in imaginative cognition (Addis & Schacter, 2012).

Arguably, the evolution of imagination led to a distinct way of being and interacting with the world. Animals that imagine have a unique experiential perspective of their environment and of their selves. This research brings to light the evolutionary origins of a neural network widely studied in humans in diverse imaginative tasks. There are reasons to think that studying this cortico-hippocampal network from an evolutionary perspective will yield important insights in the future.

Maria Sekatskaya

11.3 - TOWARD A MODESTLY PRESCRIPTIVE EVOLUTIONARY ETHICS

Maria Sekatskaya, Ruhr-Universität Bochum

Abstract:
Naturalized ethics is a promising interdisciplinary approach that seeks to integrate ethical theory with the findings of empirical science—especially evolutionary biology, psychology, and neuroscience. Among its key strands is evolutionary ethics, which investigates the evolutionary origins of moral behavior and its normative implications. While descriptive evolutionary ethics, according to Richard Joyce's influential book The Evolution of Morality (2006), is a legitimate research program and a valuable tool for explaining the emergence of morality, the status of prescriptive evolutionary ethics, which aims at deriving moral norms from biological evolutionary facts, remains controversial. This talk defends a modestly prescriptive version of evolutionary ethics: one that avoids the problems of overly simplistic attempts to derive moral conclusions from particular evolutionary facts, while still claiming that evolutionary explanations can inform and constrain ethical theorizing in productive ways.

First, I will briefly review the problems of prescriptive evolutionary ethics as presented in Joyce (2006). While some of Joyce's critiques are well-founded in the texts of some proponents of evolutionary ethics, I argue that these criticisms can be avoided by a more modest form of prescriptive evolutionary ethics. In particular, Joyce targets overly ambitious or naive forms of prescriptive evolutionary ethics—those that infer moral truth directly from evolutionary fitness and claim that moral properties just are whatever was evolutionarily advantageous. An example of such a position (admittedly a strawman example) is the claim that if moral norms have evolved in order to help one's group members and fight its enemies, this is what moral norms ultimately are and will always remain. I will consider some of the problems of prescriptive evolutionary ethics and their criticisms by Joyce and other authors, and argue that based on the criticism of these extreme positions, we open up the space for a more refined, modestly prescriptive evolutionary ethics.

After that, I will highlight some important insights from Patricia Churchland's Conscience: The Origins of Moral Intuition (2019). Churchland presents a biologically grounded account of moral psychology, emphasizing three central claims: (1) moral intuitions are rooted in neural circuitry shaped by evolution; (2) moral motivation is driven by social emotions and reward-based learning; and (3) while specific moral norms are not innate, the capacity to acquire and internalize norms is a product of evolutionary selection. While all three of these claims are already recognized to a certain extent even in Joyce (2006), I want to show that by properly developing the third claim, we can formulate the main principles for a modestly prescriptive evolutionary ethics.

The fact that humans evolved mechanisms for acquiring moral norms—rather than fixed moral content—suggests that morality is flexible, historically contingent, and culturally modifiable, but also constrained by the biological architecture of our moral psychology (cf. Haidt 2012; Henrich 2020). These constraints are not normative rules but empirical facts about

what kinds of moral cognition and affect humans are capable of. Evolutionary ethics, on this view, is not about deriving "ought" from "is," but about understanding the range of possible "oughts" that can be psychologically internalized and motivationally perceived as having normative force. This is why some rules are accepted by us as moral rules, whereas others are seen as purely conventional ones—a fact that is very important for the argument against prescriptive evolutionary ethics in Joyce (2006).

One way to make this idea precise is to examine the structure of moral motivation. Suppose we accept that moral judgment involves a characteristic emotional profile—what we might call moral feelings—combined with some sensitivity to rational considerations. This combination is supported by recent work in moral psychology, affective neuroscience, and social cognition. If this is the evolved structure of human moral psychology, then prescriptive claims must be constrained by this mechanism. We cannot expect to motivate action purely through abstract reason, nor should we assume that every moral emotion is a reliable guide to moral truth. Rather, we must work with a dual-aspect system: biologically grounded affective intuitions and culturally shaped rational modulation.

This framework allows us to avoid a common fallacy in both evolutionary ethics and its critics. The fallacy runs as follows: since a given trait evolved under particular selection pressures (e.g., cooperation within groups, hostility toward out-groups), it must remain tied to those original functions. But this confuses evolutionary origin with current function and ignores the plasticity of evolved systems. Just as the human visual system, though evolved for navigating the ancestral environment, now enables us to read books, so too the moral system—while shaped by specific evolutionary concerns of our ancestors—can be extended and redirected. However, what we can and cannot read depends on the evolved mechanism of our vision (for example, we cannot read a book in the dark or from a 100-meter distance). In the same way, what we can and cannot perceive as morally relevant depends on our evolved moral feeling.

This is where the concept of moral progress becomes relevant. If our moral feelings are responsive to certain stimuli under certain conditions—say, observable suffering or violations of fairness norms—then understanding these conditions gives us leverage. We can study the neural structures that underlie moral feelings and the stimuli that activate them. We may find, for example, that these structures reliably respond to suffering, even across unfamiliar group boundaries, when such boundaries are de-emphasized. This allows for the expansion of moral concern—a process seen in the historical broadening of human rights discourse.

Joyce warns that evolutionary ethics risks collapsing into moral realism (by reifying evolved responses) or into nihilism (by undermining moral justification altogether). But a modestly prescriptive evolutionary ethics avoids both extremes. It does not claim that certain properties are objectively moral and that our brains detect them, in the way that eyes detect wavelengths of light. Instead, it proposes that moral facts are facts about what kinds of stimuli reliably elicit moral responses in human beings under certain conditions.

Thus, a modestly prescriptive evolutionary ethics proposes the following: given that human beings have a biologically evolved capacity for norm acquisition, and that moral motivation

involves certain emotional and cognitive structures, ethical theorizing should take these facts into account. Therefore, these naturalistic facts have some prescriptive force.

Francesco Fanti Rovetta

12.1 - VARIETIES OF SELF-TO-MEMORY DYNAMICS IN DEPRESSION

Francesco Fanti Rovetta, Ruhr-University Bochum

Abstract:
Autobiographical memory and the self are closely connected. Autobiographical memories provide the information to form and maintain a stable self-model (memory-to-self). The self, in turn, modulates the construction of autobiographical memories during retrieval (self-to-memory). Past research has mainly focused on the former direction of influence. However, self-to-memory dynamics are being increasingly investigated in both neurotypical and non-neurotypical individuals. In this talk, I will contribute to this line of research in two ways. First, I will elaborate upon the self-memory system framework focusing on self-related processes involved in autobiographical memory retrieval. In particular, I distinguish between long-term and situationally activated aspects of the self, as well as between its representational and agentive aspects. Second, I will introduce a general framework for understanding self-to-memory dynamics in major depressive disorder based on four dimensions of the self: self-coherence (divided into self-coherence proper and self-congruence); self-valence; self-knowledge; and self-control. Based on the characteristics of the self in depression, I propose an outline of self-to-memory dynamics in depressed individuals in light of existing empirical evidence. Additionally, I suggest that this framework, adjusted as needed, could be employed to describe self-to-memory dynamics in other mental health conditions.

Fenja Plate-Güneş

12.2 - INTERACTION OF SYNTACTIC FUNCTION AND NEGATION DURING SENTENCE PRODUCTION

Fenja Plate-Güneş, Goethe-University Frankfurt

Jana-Elina Jordan, Goethe-University Frankfurt

Markus Bader, Goethe-University Frankfurt

Sascha Bargmann, Goethe-University Frankfurt

Gert Webelhuth, Goethe-University Frankfurt

Abstract:

Background: While the comprehension of negation has been widely studied, much less is known about how negated sentences are produced. This is surprising considering that negation can involve word-order alternations, a central topic in production research. In German, for instance, there are two ways to express an existential quantifier within the scope of negation, i.e. a sentence with the logical form ¬∃x.P (. . . , x, . . .). With surface scope, where the linear order between the negation particle nicht (’not’) and the existential indefinite ein (’a’) mirrors the logical formula, nicht and ein must contract to kein (’no’), as shown in (1).

(1) Ich I glaube, believe dass that man one kein no weiteres further Skript script austeilen distribute wird. will ‘I believe that no further script will be distributed.’

In order to keep negation and existential quantifier lexically distinct, the indefinite must appear in front of negation, resulting in a sentence with inverse scope like the one shown in (2).

(2) Ich I glaube, believe dass that man one ein a weiteres further Skript script nicht not austeilen distribute wird. will ‘I believe no futher script will be distributed.

Studies on sentence comprehension have shown a preference for surface scope over inverse scope. The former is semantically transparent and easily accessible, whereas the latter is considered more processing intensive (Wurmbrand 2018). Nevertheless, several recent studies (e.g., Radó and Bott 2018; Fanselow et al. 2022) have shown that inverse scope is accessible in German, suggesting that specific factors offset the advantages of surface scope.

Hypotheses: We investigate whether the syntactic function of the existential NP influences the choice between ein-nicht and kein. Research on sentence planning shows that subjects are typically planned early in sentence processing, before planning the verb phrase. Objects, on the other hand, are planned later on (Momma and Ferreira 2019). Given the incremental nature of sentence production, we derive the following predictions: (i) Subjects should precede negation more often than objects. (ii) Subjects with an agent role (unergative verb) are subjects both on the surface and underlyingly, while subjects with a patient/theme role (unaccusative and passive verbs) have properties of underlying objects and also behave more

like objects with regard to sentence planning (Momma and Ferreira 2019). Consequently, we expect sentences with unergative verbs to elicit more recalls with ein pre- ceding negation compared to unaccusative and passive verbs. (iii) For sentences with an indefinite object, we expect less inverse scope (ein nicht) for PP than for accusative objects because PP objects can appear after the negation without contraction.

Method: Our experiment employs a variant of the production-from-memory paradigm (Bader 2014). It follows a 1x5 factorial within-subjects design, with syntactic function of the indefinite NP as the independent variable at 5 levels: subject of unergative verb, subject of unaccusative verb, subject of passive verb, accusative object, prepositional object. Participants (n = 15 so far) read out a context sentence and the target sentence, which was a negated main clause with a sentence initial indefinite. For the sentences in (1) and (2), the corresponding main clause is shown in (3):

(3) Ein a weiteres further Skript script wird will man one nicht not austeilen. distribute
'One will not distribute a further script.'

Participants then read the context again, followed by a prompt like "Es heißt, dass …" ('It is said that …'), and inserted the target sentence. To do so, participants had to transform the main clause into an embedded clause. During this transformation, participants had the choice of placing the indefinite before or after the negation within the embedded clause. This variant of production-from memory allows us to determine under which conditions participants produce sentences with either kein or ein-nicht. In addition to the 30 experimental items, there were 60 filler items. Nine filler sentences (10% of the overall items) contained a negation with "kein".

Results: We computed a mixed-effects model, using the lme4 package (Bates et al. 2015) to test the hypotheses derived above in a preliminary analysis of the data from 15 participants. With regard to the predictions derived above, the model in Table 2 shows the following: (i) ein-nicht is produced significantly more often when the indefinite is the subject of the sentence rather than the object. (ii) Although Figure 1 shows some variation among the three subject types, these differences are not statistically significant. (iii) ein-nicht is produced significantly more often for accusative objects than for PP objects. PP objects were often produced after the negation without contraction.

Paul Engelhardt

12.3 - When default inferences override contextual information in polysemy comprehension: Explaining fallacies of equivocation

Eugen Fischer, University of East Anglia

Paul Engelhardt, University of East Anglia

Dimitra Lazaridou-Chatzigoga, University of East Anglia

Abstract:
This paper examines the first psycholinguistic explanation of fallacies of equivocation and thereby addresses three larger questions about polysemy processing: (1) How do default comprehension inferences contribute to the processing of irregular polysemes? (2) How strongly do such inferences influence polysemy comprehension? In particular, do they ever prevail over contextual information that defeats them? (3) Which properties of the linguistic stimulus and which traits of the comprehender, respectively, modulate the influence of these inferences? Three experiments combine plausibility rating tasks with eye tracking or with individual differences measures, to address these questions for polysemous verbs, which play a key role in sentence comprehension in verb-medial languages like English (Melinger & Mauner 1999; Tanenhaus & Carlson 1989), but have received little attention in psycholinguistic polysemy research.

Research in the psychology of reasoning uses fallacies as window into automatic processes underpinning reasoning. We regard verbal reasoning as ultimately grounded in automatic language processing and adapt the approach to the psychology of language: Persuasive fallacies of equivocation shine a light on how polysemy processing, including the construction of the situation model which grounds further reasoning about the situation talked about (Zwaan 2016).

Fallacies of equivocation occur when people draw inferences from premises that use the same word (say, "bank") in different meanings or senses, e.g.: "All [river] banks are next to water. NatWest is a [financial] bank. Therefore NatWest is next to water." We are interested in fallacies that (unlike this example) have actually been made by competent speaker/thinkers. We therefore consider fallacies in historically influential philosophical arguments. Influential "arguments from illusion" (review: Robertson, 1994) use polysemous appearance verbs. In their dominant intransitive sense ("The car looks small to Claire"), "look", "appear", and "seem" function as subject-raising verbs and serve to attribute to the patient (Claire) attitudes including beliefs about the agent (Looking at the car, Claire sees it is small, believes it is small, etc.) (Brogaard, 2013). The arguments of interest discuss familiar situations of non-veridical perception (e.g., distance and perspective), where no adult believes that, say, the object is as small as it appears from afar. These contexts require a subordinate "phenomenal" interpretation of the verb, which cancels the belief implication (Maud, 1986). We follow up the suggestion that the arguments rely on default belief inferences that are defeated by disambiguating context that typically precedes the verb (Fischer et al., 2021). We present a psycholinguistic explanation of these fallacious inferences: Many irregular polysemes initially

activate an internally structured unitary representation of semantic information that is then deployed to interpret utterances which use the word in different senses (e.g., Macgregor et al., 2015; Brocher et al., 2018). Where a subset of the initially activated information is relevant for interpreting a subordinate use, this use is interpreted with the Retention/Suppression Strategy (Giora, 2012): by retaining the relevant information from the initially activated representation and suppressing the contextually irrelevant information. Where the dominant sense is far more salient than all others, complete suppression of irrelevant information is impossible (Fischer & Sytsma, 2021) and the latter enters into the situation model. This account motivates three hypotheses:

H1 - Subordinate phenomenal uses of appearance verbs are interpreted with the Retention/Suppression Strategy.
H2 - Both dominant and phenomenal uses of appearance verbs trigger default belief inferences that are supported only by the dominant sense.
H3 - The default belief inferences triggered by both dominant and phenomenal uses of appearance verbs influence text comprehension; i.e., their conclusions will get integrated into the situation model.

Three studies examined H1-H3, and whether H2-H3 hold even when the verb is preceded by a disambiguating context that invites phenomenal interpretation from the start, namely, by specifying familiar non-veridical viewing conditions. A two-round norming study (N=100, N=200) identified familiar conditions of veridical perception (where, participants think, things look their true size, shape, or colour), non-veridical perception (where things look different), and 'neutral' conditions (where one cannot tell whether or not things look their actual size, etc.).

Two eye-tracking experiments implemented the psycholinguistic cancellation paradigm. Participants read three-sentence items and rated their plausibility. In a within-subjects 2□2 design, we manipulated veridicality in the first sentence and consistency with the hypothesised belief inference ('small' vs 'large' below) in the third. Both studies examined inferences after non-veridical contexts (as in Table 1). Experiment 1 (N=45) added items with veridical contexts, Experiment 2 (N=48) added items specifying 'neutral' conditions (48 critical and 48 filler items in both studies). We measured fixation times in five regions.

Table 1. Regions of Interest
The car in the valley was far away[1]. It looked[2] small[3] to Claire[4]. She believed it was large[5].
[1]Pre-verbal context [2]Source verb [3]Source adjective [4]Source object [5]Conflict adjective

H1 predicts higher rereading times on the source verb, when this is given a phenomenal interpretation. H2 predicts higher rereading times in the inconsistent than the consistent condition for the source and conflict regions. H3 predicts lower plausibility ratings for items in the inconsistent than consistent conditions.

To assess H1, we compared rereading times between participants who provided evidence of having won through to a purely phenomenal interpretation by giving higher plausibility ratings

for inconsistent than consistent items in the non-veridical condition. In Experiment 1, too few participants provided such evidence. In Experiment 2, we found the predicted main effect of group precisely at the source verb: t = -2.16, p < .05 and the source adjective: t = -2.24, p < .05) that attracts spill-over.

Rereading times supported H2, with the predicted main effects of consistency (INCON □ CON) for both source verb (Exp.1: t =3.32, p =.002; Exp.2: t = 1.83, p = .07, but significant for source adjective t = 3.11, p = .002 and source object t = -2.40, p = .02, arguably due to spill-over) and for conflict adjective (Exp.1: t = 2.94, p = .005; Exp.2: t = 2.36, p = .02).

Plausibility ratings showed a context by consistency interaction. In Exp.1, mean ratings were INCON < CON, for both veridical and non-veridical items, as predicted by H3. In Exp.2, however, this difference was observed only for neutral, but not for non-veridical items. Since Exp.1 and 2 used the same non-veridical items, we infer that the particularly difficult neutral items acted as reflection prompts promoting deeper analytic processing (Alter et al., 2013) of all items. The pattern of total dwell times supported this interpretation. We infer that (H3*) in this setting more reflective participants will manage to suppress default belief inferences, while unreflective participants will not.

Experiment 3 (N=99) combined the plausibility rating task from Experiment 2 with individual differences measures (Need for Cognition NCS-18, Digit Span, Cognitive Reflection CRT-2, Stroop), to assess H3* and H2. Exploratory factor analysis revealed that NCS-18 and Digit span responses loaded on one factor, whereas CRT-2 and Stroop measures loaded on another. H2 predicts correlations between participants' factor scores for this "reflectiveness- and-inhibition" factor and ratings for non-veridical items (negative for consistent, positive for inconsistent items). H3* predicts that "unreflective" participants who score low on this factor will rate consistent items more plausible than inconsistent items, in both neutral and non-veridical conditions, whereas "reflective" participants with high factor scores will do so in the neutral, but not the non-veridical condition. We observed the correlations predicted by H2 (consistent: -.22*; inconsistent: .27**, whole sample; -.33* and .42**, high performers) and observed the pattern predicted by H3* when splitting the sample evenly into halves with higher and lower scores on the "reflectiveness-and-inhibition" factor.

We conclude: (1) Polysemous appearance verbs are processed with the Retention/Suppression strategy. (2) Default belief inferences from these verbs strongly influence comprehension and can prevail even over preceding disambiguating context that defeats them. (3) These inferences' influence can be reduced by stimuli that prompt reflection and by comprehender's reflectiveness.

Mira Schwarz

13.1 - Influence of Negative and Positive Ambient Scent On Wayfinding Performance and Implicit Memory

Mira Schwarz, Justus Liebig University

Abstract:
This study investigates the influence of unconsciously perceived ambient scents on human wayfinding performance. Challenging the conventional view of human navigation as primarily vision-based, this research delves into the underappreciated role of olfaction in spatial orientation and memory. Forty-one participants were tasked with learning a route through a virtual maze, utilizing either visual or mental olfactory landmarks, within test cabins subtly scented with either butyric acid (negative associated), lavender (positive associated), or left unscented as a control. These scents were administered at weak concentrations intended to remain below the threshold of conscious detection for most participants. The participants' route memory was rigorously assessed immediately following the learning phase and again after a one-month interval to evaluate the long-term retention and forgetting dynamics.

Beyond the wayfinding task, the study incorporated an implicit odor memory task first introduced and tested by Degel and Köster (1998, 1999) in order to assess whether they perceived the ambient scent in the test cabins unconsciously, consciously or not at all. Participants were presented with seven distinct odors, including the ambient scents of butyric acid and lavender, and asked to rate the perceived fit of each odor within ten different surroundings, one of which was the test cabin they had occupied. Furthermore, they were asked to label the presented odors and recall any specific instances of when and where they had previously encountered them. Participants who remembered smelling lavender or butyric acid in the test cabins were excluded from the analyses, ensuring that the observed effects were indeed attributable to implicit odor perception.

The results of the study provided compelling evidence for the implicit processing of ambient scents. The fit ratings revealed a significant increase in the perceived fit between the visual context of the test cabin and the respective ambient odor in the scented conditions compared to the no-scent condition, while the remaining participants could not report any conscious recollection of them. This finding strongly suggests that even without conscious awareness, the ambient scents were being processed and integrated with the contextual information of the environment. Notably, the ambient scent of butyric acid was found to significantly improve wayfinding performance, indicating a potential facilitative effect of this particular odor on spatial learning and navigation. In contrast, the presence of ambient lavender did not yield any significant impact on wayfinding performance. No clear evidence was found for the effects of ambient scents on forgetting dynamics over a one-month period. Interestingly, the study also uncovered interference effects between mental olfactory imagery and the implicit perception of real ambient odors. This was reflected in a reduced wayfinding performance among participants who utilized mental olfactory landmarks for navigation when ambient scents were present, compared to those in the no-scent condition. This suggests a potential

cognitive conflict or competition between internally generated olfactory cues and the unconsciously perceived ambient olfactory information.

These findings collectively underscore the critical, albeit often unconscious, role of olfaction in human wayfinding. They contribute to a growing body of research highlighting the multisensory nature of spatial cognition and challenge the traditional emphasis solely on visual cues. The observed improvement in wayfinding performance with butyric acid, despite its typically negative valence, warrants further investigation into the specific mechanisms and potential arousal effects of different odors on cognitive processes related to navigation (Yerkes & Dodson, 1908). The interference observed between mental olfactory imagery and real ambient scents also opens new avenues for exploring the interplay between internally and externally generated olfactory information in spatial tasks. The implications of these findings extend to various applied settings, including the design of environments to enhance navigation and the potential use of subtle olfactory cues in assistive technologies. Future research should delve deeper into the specific neural mechanisms underlying these effects, explore the impact of different odor concentrations and valences, and investigate individual differences in olfactory sensitivity and processing in the context of wayfinding. The study's findings emphasize the need to consider the often-overlooked sense of smell in our understanding of human spatial behavior and cognition (Schwarz et al., 2024).

Christian O. Scholz

13.2 - Aphantasia and Unconscious Imagery: A Critical Assessment from the Perspective of Shared Representations

Christian O. Scholz, Ruhr-Universität Bochum

Abstract:
For the majority of us, acts of visualization (i.e., seeing with the 'mind's eye') are a common and familiar aspect of our everyday cognition. However, an estimated 4 percent of the general population has aphantasia (Dance et al. 2022), a recently coined cognitive norm variant characterized by a severe diminution or complete absence of (visual) mental imagery (Zeman et al. 2025). Throughout the last decade, a growing body of evidence has shown that aphantasics (people with aphantasia), despite their reported lack of (visual) mental imagery, can perform a range of tasks that were previously assumed to rely on/test for mental imagery, including mental rotation (Kay et al. 2024), visual knowledge (Liu & Bartolomeo 2023) and visual memory tasks (Bainbridge et al. 2021). These findings pose a puzzle: How can people who supposedly lack visual imagery (i.e., aphantasics) solve visual imagery tasks?

Some (Liu & Bartolomeo 2023; Scholz 2024) have argued that aphantasics solve the tasks in question by employing alternative non-imagery-involving strategies. However, recently, Nanay (2021, 2023) proposed that the puzzle can be solved by positing that aphantasics have unconscious mental imagery, meaning that despite not experiencing mental imagery, they may still possess the underlying neural correlate associated with mental imagery (see also Michel et al., 2025 for a recent defense of this position).

But what are the empirical criteria for assessing (unconscious) mental imagery on Nanay's account? Nanay, following psychological accounts (Pearson et al. 2015; Kosslyn et al. 2006), defines mental imagery as "perceptual representation that is not directly triggered by sensory input" (Nanay 2023, p. 4). Importantly, he explicitly states that both perception and mental imagery are instances of perceptual (neural) representations, with the "mark of mental imagery" (Nanay, 2023, p. 9) being the absence of a direct link between sensory input and (neural) representation. Thus, since, on Nanay's view, the representations in imagery and perception can be merely distinguished by the way they are triggered, I posit that imagery and perception, for him, must utilize shared representations (Scholz et al., 2025).

One way of assessing whether two cognitive processes (e.g., perception and imagery) rely on shared representations, is to assess their cross-decodability, by first training machine learning algorithms on the activity pattern derived from one condition (e.g., perception) and then testing whether said algorithms can decode the content of patterns obtained in the other condition (e.g., imagery), and vice versa (see Naseralis et al., 2015 for evidence showing cross-decodability between perception and mental imagery in typical visualizers). A methodologically related technique is to compute the so-called representational overlap (Kriegeskorte et al. 2008), by assessing the structural similarities between the representations observed in two conditions, in order to identify and quantify shared representations. Since Nanay's definition of mental imagery is built on the claim that mental imagery uses the same

type of representation as perception does (i.e., “perceptual representation”), I thus posit that if aphantasics have unconscious mental imagery during imagery tasks, then their neural activity must rely on shared (perceptual) representations, as assessed via cross-decodability and representational overlap.

However, two recent studies (Chang et al. 2025; Liu et al. 2025) provide trouble for the proposal that aphantasics rely on unconscious mental imagery. In both studies, despite aphantasics showing stimulus-specific activation in the visual cortex, the observed activity did not rely on shared representations. Importantly, while one of the studies (Chang et al. 2025) investigated cross-decodability in early visual areas, the other study (Liu et al. 2025) focused on representational overlap in higher visual areas, with both of them reporting finding shared representations only in the non-aphantasic control group. Thus, shared representations were neither found in early nor in higher visual areas in aphantasic subjects during imagery tasks.

Taken together, these results generate tension for Nanay’s definition of mental imagery. On the one hand, the results indicate that aphantasics have activity in visual areas and that this activity is representational, in the sense that it is stimulus-specific. However, on the other hand, since these representations do not utilize shared representations, they should not be viewed as perceptual representations, as posited by Nanay. Thus, Nanay and the proponents of the unconscious imagery view, have to either a) give up on the idea that aphantasics rely on unconscious imagery, or b) give up on the definition of mental imagery as involving perceptual representation. However, if they opt for option b), then it is not at all clear what the concrete empirical criteria for assessing whether aphantasics possess unconscious imagery or not are supposed to be. For, if one merely requires aphantasics to have stimulus specific representations in perceptual areas, there is no guarantee that these representations are indicators for unconscious imagery rather than alternative cognitive or representational strategies. For example, a range of findings show that the congenitally blind often show activity in early visual areas (Bedny, 2017), though these representations are, due to neural plasticity, rather associated with non-visual, for example, haptic representations. Similarly, Anderson’s (2010, 2014) concept of neural reuse impressively shows the variety of functions and abilities that can rely on functionally distinct activity and connectivity in the same brain regions.

In conclusion, despite finding stimulus-specific representations in aphantasics’ visual brain regions, I argue that we should be cautious to not interpret these as evidence for the claim that aphantasics have unconscious mental imagery. For, if proponents of the unconscious imagery claim follow the proposed definition of mental imagery as perceptual representations, this forces them to accept that unconscious imagery relies on shared (perceptual) representations. However, the best available studies on shared representations in aphantasics, show that aphantasics do not exhibit shared representations during imagery attempts. Therefore, Nanay (and other proponents of unconscious imagery) must either abandon their claim that aphantasics have unconscious mental imagery, or propose new empirical criteria supposed to characterize mental imagery representations.

Jonida Kodra

13.3 - CONTROL IN AUDITORY VERBAL HALLUCINATIONS

Jonida Kodra, University of Osnabrueck

Abstract:

Auditory verbal hallucinations (AVHs) are broadly defined as the experience of hearing voices in the absence of an appropriate stimulus (APA 2022; David 2004; Slade & Bentall 1988). Although they are mostly known for being one of the main symptoms of schizophrenia, they are also common in other disorders and in the healthy population (Beumeister 2017). AVHs are not just clinically relevant, but also philosophically important because they raise crucial questions about the nature of the experience (Langland-Hassan 2018), mental agency, and delusional beliefs (Stephens & Graham 2000). Despite the wide body of research, many aspects of AVHs are still unclear. One such aspects is control. Understanding control in AVHs is important for at least three reasons. The first is that control is considered one of the aspects that distinguishes pathological from non-pathological AVHs experiences (Murphy 1976; Beumeister 2017). The second is that a lack of control over AVHs is taken to indicate that AVHs are non-veridical perceptual as opposed to cognition-like experiences. Indeed, philosophers (Farkas 2012; Knappik et al., 2022) who distinguish between those AVHs that are subjectively indistinguishable from veridical auditory perceptions and those AVHs that are subjectively distinguishable from them, maintain that a lack of control differentiates AVHs that are subjectively distinguishable from other inner experiences such as mental imagery. The third is that control is either identified with or assumed to be closely tied to what the clinical and philosophical literature refers to as sense of agency. A lack or a disrupted sense of agency is commonly taken to be the reason why voice-hearers cannot recognise their AVH episode as self-generated (Waters et al. 2006; Wu 2012).

In general, all clinical definitions agree that voice-hearers lack control, but they disagree on whether what they lack is the feeling or sense of control (Lingiardi & Williams 2017; David 2004), or actual control (APA 2022; Slade & Bentall 1988). Clinical definitions are also at odds with both psychological assessment tools and empirical studies, the former assuming in their questionnaires that voice-hearers can have control, and the latter showing that this is indeed the case.

The topic of control in AVHs is largely ignored in more conceptual work, and only been the focus of one recent review of the literature by Swyer & Powers (2020). This review highlights that the clinical literature understands control as actual control, that is, as “an ability to voluntarily influence voice-hearing experience” (p. 2). In the attempt to clarify the clinical literature around control, they divide it into two broad categories reflecting, according to them, two distinct types of control. Their distinction, however, does not pick out different types of control, but rather one type of control – namely the agent’s ability to initiate or inhibit AVHs episodes –, one which can be achieved through strategies that involve either direct control, or indirect control. The clinical literature has also shown that voice-hearers often report other forms of control: some voice-hearers indeed reports that they can change some features of their AVH experiences (Nayani & David, 1996); others, that they feel like they are controlled by the voices (Chadwick et al., 2000; Romme & Escher, 1993). Given the current incomplete notion of control, I suggest that none of the philosophical issues mentioned can

be answered. A more thorough and detailed conceptualization of control – one that is based on a more careful consideration of how the experience of control is reported by voice-hearers – is thus needed.

Specifically, the suggestion is that, to clearly and properly understand control, we need to distinguish two key aspects: first, who is exercising control and, second, what control is being exercised. Concerning the who, every AVH episode involves at least two agents: one is an actual agent, the voice hearer himself, and the other is an imagined one, that is, the agent that the voice-hearers reports that the voices belong to. Concerning the what, the empirical literature points out that both agents seem to exercise some form of control. Voice-hearers exercise actual control, which is their ability to affect the AVH episode or some of its features or to refuse to act in accordance: I call this active and passive actual control. Voices, on the other hand, can be thought of as exercising control resulting in a sense of being controlled: this is experienced control. From the actual agent's perspective, this involves the feeling that the voices have the ability to affect the episode by either initiating or inhibiting it, by influencing voice-hearer's thoughts or actions, or by forcing the voice-hearer to act on commands. I call this active and passive lack of sense of control.

This way of conceptualising control presents two main advantages. The first is that it helps, together with the empirical data, towards clarifying the three assumptions I mentioned earlier. The second is that it will also help us, in general, to better understand the notion of mental agency more generally, namely, how sense of subjectivity, control, and sense of agency relate to one another. In more recent literature, philosophers have distinguished between what is called sense of subjectivity and sense of agency (Stephens & Graham, 2000). When one has a sense of subjectivity, one is introspectively aware of some mental event; when one has a sense of agency, on the other hand, one has the feeling of being the author or agent of a mental event, of having brought it about. The distinction between these experiences lies on the idea that thinking is an action. This means that, when saying, for instance "The voice in my head is not mine," voice-hearers mean that they did not generate that episode. Voice-hearers report that, although they are able to exert some control over their AVH episode, they still feel like they have not generated the episode (Nayani & David, 1996; Romme & Escher, 1993). This seems to suggest that control is not an essential feature of mental agency.

Aalia Nosheen & Annette Hohenberger

14.1 - A Situational Judgement Test to Measure Critical versus Dialectical Thinking: Examining the Roles of Culture and Gender

Aalia Nosheen, University of Osnabrueck

Annette Hohenberger, University of Osnabrueck

Abstract:
While critical thinking is widely known and appreciated in higher education, dialectical thinking is less generally known. Critical thinking concentrates on the problem analytically, identifies its direct causes, and assumes long-term stability. In contrast, dialectical thinking considers the whole context of a problem, including direct and indirect causes, assumes only temporary stability, and aims to achieve a balance between opposites (Spencer-Rodgers et al., 2018). So far, critical thinking has been considered as a skill, measured by objective tests, but dialectical thinking only as a thinking tendency, measured by subjective questionnaires. Whereas critical thinking tests are commercial products and their use is restricted, dialectical scales are free of charge. Here, we developed a situational judgment test to measure critical versus dialectical thinking with the same objective instrument for the first time. In accordance with the principles of open science, the test will be made publicly available as a psychometric assessment and research tool.

Critical thinking is often aligned with "Western," "individualistic" culture and dialectical thinking with "Eastern," "collectivistic" culture (Nisbett, 2003; Rear, 2017). The little research that exists on the role of gender in both types of thinking has produced inconsistent results so far (Kramer & Melchior, 1990; Bagheri & Ghanizadah, 2016). Therefore, we examined the role of culture and gender in critical and dialectical thinking in two samples comprising male and female students, from Germany and from Pakistan, respectively.

The situational judgement test was developed according to the principles of psychometric test construction and advised by experts from both cultures. It was designed in English—the language of study of both the German and the Pakistani students. The test comprises 27 items consisting of short scenarios followed by three response options: a critical, a dialectical, and a distractor response. The scenarios measure six skills that comprise higher-order thinking (evaluation, self-regulation, inference, explanation, interpretation, and analytical skills) (Facione, 1990). However, the first five items of this test measure only critical thinking (versus the distractor option) using mathematical, probabilistic problems, whereas the remaining 22 items measure critical versus dialectical thinking (versus the distractor option) through commonly experienced situations.

First, a pilot study was conducted. It assessed agreement strength for each response option separately via Likert scales. This allowed us to study endorsement rates of critical, dialectical, and distractor responses independent of each other. Results showed that mean scores of the German sample were higher on critical thinking as compared to the Pakistani sample, whereas mean scores of the Pakistani sample were higher on the distractor response. Both cultures endorsed dialectical thinking to the same extent. Based on feedback from participants in the

pilot study, some test items were revised. Subsequently, the test was conducted in the choice format explained above with (n=123) students from Germany and (n=109) students from Pakistan.

In order to unravel the factor structure and content validity of our situational judgement test, we used latent class analysis (LCA). Findings strongly suggest a 2-class model as the best fit, with a high probability for critical and dialectical responses but only a low probability for distractors. Thus, LCA indicates that participants from both cultures employ both critical and dialectical thinking skills as forms of higher-order cognition. Convergent validity was assessed by examining how well the newly constructed test corresponded with previously standardized scales. It demonstrated good convergent validity with two subjective scales measuring dialectical thinking, i.e., the Analysis Holism Scale (Martín-Fernández et al., 2022) and the Holistic Cognitive Scale (Lux et al., 2021), across cultures. Additionally, our test showed good convergent validity with the Triad Task (Li & Hu, 2022), assessing categorical versus relative judgement, within each culture.

Considering the role of culture and gender, significant differences were found in the situational judgement test. German participants scored higher on both critical and dialectical thinking as compared to Pakistani participants, who endorsed distractors more often in their choices (see Fig. 1). Moreover, participants in both cultures unanimously preferred dialectical over critical thinking, indicating a possible global shift in higher- order cognition over the last decades, as compared to earlier studies (Nisbett, 2003; Eigenauer, 2006; Rear, 2017).

Considering gender, overall, male participants endorsed critical thinking options more, whereas female participants endorsed dialectical thinking more. However, a cross- cultural comparison of gender showed that the gender difference was significant in the German culture only, but not in the Pakistani culture. This finding may indicate that in individualistic cultures, gender differences in cognition emerge more easily, as compared to collectivistic cultures, where both genders have to manage complex social and kin relationships that demand similar thinking styles, irrespective of gender (Guimond, 2008).

In addition, the current study found that participants from both cultures showed lower performance in the first five items (see again Fig. 1), which demanded some mathematical or probability calculations, i.e., they chose more distractors, as compared to the other 22 items, which tapped critical and dialectical thinking in daily life situations. This finding is in line with earlier results showing that—even in higher education— students would rather employ cognitive heuristics and biases than logic and probability calculus (Battersby, 2016).

The results of our current study on critical and dialectical thinking across cultures and gender suggest that it is time to rethink long-standing but possibly outdated beliefs about their nature and use in people's reasoning. One promising way to overcome the previously held dichotomy between critical and dialectical thinking is construing them in terms of "metathinking" (Shannon & Frischherz, 2020). Metathinking is the ability to evaluate one's own thinking process and to choose flexibly to use either, depending on the context. Acknowledging both types of thinking as equally valid instantiations of higher-order thinking, we suggest that

students of different cultural backgrounds may reason flexibly in response to short- and long-term changes in their physical, social, political, and educational environment. We believe that metathinking captures the requirements of necessary cognitive adaptations of the young generation to a changing, ever more complex world at local and global scales.

Evelyn Ferstl

14.2 - GENDER-FAIR LANGUAGE BEYOND ORTHOGRAPHIC NORMS: EYE-TRACKING EVIDENCE ON THE READABILITY OF THE BINARY CAPITAL-I FORM AND THE NONBINARY GENDER STAR IN GERMAN

Lisa Zacharski, University of Freiburg, Leibniz Institute for Prevention Research and Epidemiology - BIPS

Annika Oldach, University of Freiburg

Evelyn C. Ferstl, University of Freiburg

Abstract:
The common traffic sign “Radfahrer bitte absteigen” (Cyclists [masc.] please dismount) illustrates the generic use of the masculine form in German. While the feminine form Radfahrerinnen refers only to female cyclists, the masculine Radfahrer may refer specifically to male referents, or generically to both women and men, if gender is unknown or irrelevant (cf. Gygax et al., 2019). However, numerous experimental studies have shown that the so-called generic masculine leads to a substantial male bias (cf. Gygax et al., 2021). To address this, beginning in the 1980s, binary alternatives that aim to increase the visibility of women were suggested, such as the capital-I form (Binnen-I, e.g., RadfahrerInnen – feMale cyclists). However, binary forms have been criticised more recently for excluding persons identifying beyond the female-male dichotomy, which has led to the development of nonbinary forms (cf. Diewald & Steinhauer, 2017). The form that has gained significant traction over the last couple of years (cf. Krome, 2020; Meuleneers, 2024) is the gender star form (Genderstern; Radfahrer*innen). Initial studies suggest that it reduces male bias and prompts all-inclusive mental representations (Zacharski & Ferstl, 2023). Despite this, the German Council for Orthography (Rat für deutsche Rechtschreibung, 2021) opposes all gender-fair forms that are not in line with German orthography (from now on: non-orthographic forms), arguing they negatively affect readability and comprehensibility. However, these claims have yet to be thoroughly tested experimentally.

For binary non-orthographic forms, research suggests that their use leads to worse subjective ratings of readability (Pöschko & Prieler, 2018) and longer reading times (Blake & Klimmt, 2010) compared to the masculine form. For the nonbinary gender star, Friedrich et al. (2021) found that subjective ratings of comprehensibility in a student sample were worse only when the text contained predominantly singular forms, which require adaptations of articles and adjectives (e.g. der*die Spieler*in – the[masc.]*the[fem.] player[star]). Pabst & Kollmeyer (2023) found no evidence that the gender star reduces the comprehensibility of texts as measured with subjective ratings, neither in an academic and non-academic adult sample. While these studies offer valuable insights, subjective measures are susceptible to metacognitive influences like the social desirability bias. Therefore, experimental studies with more implicit measures are needed.

One such study (Zacharski et al., 2024, in press) used a lexical decision task to examine word-level processing. Interestingly, reaction times for recognition of gender star role nouns were

as fast as for masculine ones in a young student sample. In contrast, an older non- academic group initially showed slower recognition, possibly due to less familiarity, though this effect diminished over time. Eye-tracking is another implicit method, offering detailed, real-time data on cognitive processing during reading, by recording eye-movements (cf. Rayner & Sereno, 1994; Schotter & Dillon, 2025). It provides objective measures of processing effort for pre-determined areas of interest (AOIs) within a text. The present study is the first to use it to examine the readability of two non-orthographic gender forms, the binary capital-I form and the nonbinary gender star, at the text level, with both a young student sample (Experiment 1), and an older, non-student sample (Experiment 2).

We used a 3x4 within-subjects eye-tracking design to examine how gender form (masculine: e.g., Mieter vs. capital-I: e.g., MieterInnen vs. gender star: Mieter*innen – tenants) and the position of the role noun within the text (position 1–4) impact readability, comprehensibility, and memory of texts. Participants read 32 short, newspaper-style texts (~120 words) covering different topics (example text see Figure 1), including 24 experimental texts featuring four instances (i.e., tokens) of a stereotypically neutral (Misersky et al., 2014) plural role noun. Gender form was varied across participants, with each individual reading eight texts per form. Additionally, eight filler texts contained only neutral alternatives (e.g., participles: Geflüchtete – refugee). To mask the manipulation, each text depicted a custom- designed logo of a fictional regional newspaper or blog, thematically matching the article content. Following each text, participants answered one multiple-choice comprehension question. After the eye-tracking session, they completed 24 free-text recall questions, testing their memory for the role nouns. Finally, participants completed a questionnaire on their attitudes towards binary and nonbinary gender-fair language (ABNBL; Zacharski, 2024) and provided demographic information.

Based on previous findings by Zacharski et al. (2024, in press), we expected interactions between gender forms and participant characteristics: In Experiment 1, using a young student sample (<30 years of age) with overall positive attitudes towards gender-fair language and nonbinary forms in particular, we expected the gender star form to reduce processing effort, compared to the capital-I form. However, we anticipated increased processing effort for the gender star form in contrast to the masculine form, due to its non- orthographic character. In Experiment 2, we use an older, non-student sample (40–70 years of age), who are likely more familiar with the previously popular capital-I form and are expected to hold less positive attitudes towards gender-fair language, particularly nonbinary forms (cf. Zacharski, 2024). Hence, we hypothesize that the gender star will increase processing effort, compared to the masculine, while this effect is less pronounced for the capital-I form.

As data collection for Experiment 1 is still ongoing and data collection for Experiment 2 is about to begin, we conducted preliminary analyses for Experiment 1 based on data from 38 psychology or cognitive science students (f: 26, m: 12). These analyses employed a two-step linear mixed model approach, with role nouns defined as AOIs. First, variance caused by known predictors of reading behaviour (e.g., word length) was factored out. In the second step, the residuals were modelled as a function of gender form and word position, with trial number, role noun frequency, and participants' attitudes towards gender-fair language

included as covariates. Four eye-tracking metrics known to be indicative of linguistic processing effort (cf. Rayner & Sereno, 1994; Schotter & Dillon, 2025)—namely, first fixation duration, first pass reading time, total reading time, and regression path duration—were used as dependent variables. Comprehension and recall data were analysed via generalized mixed-effects models.

As hypothesised, first pass reading times were significantly faster for gender star role nouns, compared to the capital-I form, suggesting easier processing for this sample (see Figure 2). This finding aligns with the familiarity effect described by Zacharski et al. (2024, in press), as the gender star form is particularly popular among this demographic. Surprisingly, however, reading times were also significantly faster for gender star role nouns compared to the orthographic masculine form, suggesting that the gender star not only does not impede readability but may even enhance it. No significant differences were found between capital-I and masculine forms, or in comprehension or recall accuracy across gender forms.

At the conference, we will present findings from the final sample of Experiment 1 and preliminary results from Experiment 2. Comparing both samples will provide insights into familiarity and attitude effects. In particular, we aim to determine whether the reduced reading times for the gender star in Experiment 1 are mirrored for the capital-I form in Experiment 2, or if both non-orthographic forms increase reading times. Results will be discussed considering reading theories and language policy.

Asya Achimova

14.3 - THE ELECTION OUTCOME WAS INTERESTING! STRATEGIC INDIRECTNESS IN A CROSS-CULTURAL PERSPECTIVE

Asya Achimova, University of Tübingen

Dirk Wildgruber, University of Tübingen

Martin V. Butz, University of Tübingen

Abstract:
One of the virtues of efficient communication, according to classic rhetoricians (Ossa-Richardson, 2019), is ambiguity avoidance. Intuitively, ambiguous words and utterances may be confusing for the listener and choosing such utterances may violate the maxims of cooperative communication (Grice, 1975). These maxims prescribe speakers to choose most efficient ways to communicate intended messages. Ambiguous messages fall short on the efficiency metric because they appear suboptimal from the point of view of information transfer: the listener may be uncertain which of the meanings of an ambiguous message is actually communicated.

However, ambiguity may carry social benefits. In politeness contexts, it offers speakers an opportunity to remain non-committal to an utterance if its contents turns out to be problematic (Gotzner & Scontras, 2024). A notable property of polite forms is their indirectness (Yoon et al., 2020). The meaning of indirect utterances is albeit compatible with the true opinion of the speaker but it does not communicate it in the most straightforward way. Indirect communication is not limited to politeness situations: it can be used strategically to communicate content that is socially sanctioned, such as offering an officer a bribe, as in (1).

(1) Gee, officer, is there some way we could take care of the ticket here? (Pinker et al., 2008, p. 833)

Indirect communication may also offer an escape strategy to well-meaning speakers who use indirect utterance as a way to avoid a possible conflict of opinions with the listener (Achimova et al., 2023). A speaker who is very excited about an election outcome may prefer a more subtle predicate “interesting” as in (2) if she anticipate that her opinion may not align with the opinion of the listener. A subsequent reaction of the listener may reveal her beliefs and give an idea to the speaker whether their beliefs align. Both in (3-a) and in (3-b) the speaker expresses an agreement with the listener but (3-a) shows that the listener interpreted “interesting” as a positive evaluation, while (3-b) reveals a negative interpretation of the same predicate.

(2) The election outcome was interesting!
(3) a. Yes, such a relief!
b. Yes, a pure disaster!

The main challenge for formal models of indirect communication is to explain why these utterances are an optimal speaker choice given that they are ambiguous. Achimova et al. (2023) develop a probabilistic model of utterance choice where indirect utterances allow the speakers to simultaneously achieve informational and social goal of avoiding potential conflict.

In this work, we investigate whether the search for belief alignment is a universal strategy speakers use or whether it is culturally modulated. The alignment model of indirect communication predicts that speakers will be more likely to choose indirect utterances when they anticipate a possible belief clash with the listener (Achimova et al., 2023). However, the model has only been evaluated with US-English speakers and it remains unclear whether this behavior generalizes. Moreover, in the aforementioned work, participants evaluated dialogues of simulated speakers which might not necessarily translate to their own communicative behavior. To address these shortcomings, we designed a study that compares the choice of utterances by English and German speakers in situations where they need to discuss controversial topics with a simulated conversation partner. Below we report preliminary data and a qualitative analysis of these results.

Sebastian Scholz

15.1 - CONCEPTUAL SPACES AS PREDICTIVE MODELS?

Sebastian Scholz, Heinrich-Heine-Universität Düsseldorf

Abstract:
Conceptual Spaces (CS) and Predictive Processing (PP) are prominent frameworks in cognitive science, yet their relationship remains largely unexplored. CS models conceptual structure through similarity, while PP views cognition as probabilistic (or Bayesian) inference aimed at minimizing prediction error. Recent work shows how CS can generate priors for Bayesian reasoning, and how probability distributions can be modeled on CS, bridging these frameworks in ways that reveal their complementarity and shed new light on the interplay between (psychological) similarity and probability.

This interplay was extensively discussed in the mid-20th-century. Many psychologists of that era were wary of systematic errors of judgment which had cast doubt on the reliability of the human mind and its presumed superiority over non-human animals. As Schurz & Hertwig (2019, p. 8) point out, such worries were based on a single widespread methodological assumption: "Either individuals behave in accord with the chosen benchmark of rationality, or their cognitive behavior, measured against the benchmark, is irrational [….]." Most prominently, the heuristics and biases program (Tversky & Kahneman 1974) treated similarity-based reasoning as an efficient, but ultimately irrational shortcut for probabilistic reasoning, which was seen as the benchmark of rationality. Today's theorists tend to treat both kinds of reasoning as potentially rational, or as adaptive depending on the available cognitive capacities and reasoning environments at hand (Gigerenzer & Todd 1999, Sloman 1996). In other words: they are seen as complementary aspects of reasoning, or cognition more generally.

I come to the same conclusion, but on grounds that go beyond the heuristics and biases debate. First, note that there are strong prima facie reasons to assume a high level of compatibility between CS and PP. The latter underlines the dynamic and interactive nature of cognition, while interaction lies at the heart of the most prominent version of CS as well: "Via successful and less successful interactions with the world, the conceptual structure of an individual will adapt to the structure of reality" (Gärdenfors 2000: 156). Given the diversity of PP accounts, however, two compatibility constraints must be in place: openness to non-probabilistic representations and descriptivism (CS is a descriptive account of cognition, while PP accounts are often depicted as normative – but they also aim to describe cognition in a "grand unified theory"). When these conditions are met, CS and PP can be seen as operating on different levels of cognitive processing in Marr's framework (cf. Marr & Poggio 1976). For instance, PP is sometimes discussed at the computational level, as it specifies what cognition is ultimately about: minimizing prediction error by probabilistically inferring the causes of sensory input. This is a broad, abstract goal and doesn't specify the algorithms or mechanisms used to achieve it – which can make these models difficult to interpret semantically. CS, on the other hand, is most naturally understood as operating on the algorithmic level. It provides a framework for structuring and manipulating conceptual representations, such as through

similarity metrics, prototypes, and dimensions. These entities have not only mathematical, but cognitive content and are derived from experience. Thus, CS models specify the algorithmic tools the mind uses to carry out cognitive tasks like categorization and reasoning, potentially acting as a way of achieving the computational goals of PP in a way that is semantically interpretable.

While these considerations pertain mostly to the "division of labor" between the two frameworks, recent advances show how CS models can contribute to probabilistic reasoning and offer concrete mechanisms for integrating the two frameworks horizontally on the algorithmic level. Decock et al. (2016) utilize the properties of conceptual regions, such as their relative size, to generate priors for Bayesian updating. Recent extensions of this approach by Douven et al. (forthcoming) incorporate iterative belief updating, aligning well with PP's dynamic, interactive view of cognition. They define a single probability distribution on a Conceptual Space, whereas Sznajder (2017) defines a distribution over a set of hypotheses about distributions in CS, which is more in line with traditional Bayesian models and PP accounts. I will review this literature and address open questions, including considerations about the ontology of the models and whether other geometric methods, such as distance to prototypes, could generate meaningful priors from CS.

Finally, I resist reductionist accounts, such as Tenenbaum and Griffiths' (2001) claim that their Bayesian model of perceptual categorization is fundamental because it does not depend on any specific account of similarity. Instead, I argue that similarity is crucial for the semantic interpretability of Bayesian models and PP, a point underscored by both the heuristics and biases debate and contemporary critiques (cf. Poth 2019).

In sum: If PP accounts meet certain constraints, they are generally compatible with CS. The Marr-Analysis is superior to reductionism and shows how the mind can be both a similarity machine and a probability engine. Most importantly, the work by Decock, Douven, and their collaborators demonstrates how these mechanisms interlock in the human mind, and how CS can function as predictive models.

Jannis Friedrich

15.2 - Enriching Predictive Processing to account for Higher Cognition: Simulation, Language, and Conceptual Spaces

Jannis Friedrich, Department of Performance Psychology, German Sport University Cologne

Martin H. Fischer, Potsdam Embodied Cognition Group, University of Potsdam

Abstract:
Predictive processing (PP) has positioned itself as a unifying framework in cognitive science, positing that perception, action, and cognition all rest on the mechanism of prediction error minimization (Clark, 2015; Hohwy, 2020). Despite PP's success in accounting for many diverse phenomena, one challenge persists: PP struggles to explain how this mechanism could manage to account for the representation of abstract concepts, and therefore higher-level cognition generally. This paper posits a solution to this issue by looking to embodied theories of mental representation, and explains how the basic machinery of PP could account for even higher-level abstract cognition. To this aim, we synthesize literatures on PP and Grounded Cognition (GC) and offer a coherent integrative account of higher- level cognition. Furthermore, we embed PP into biogenic approaches to cognition, demonstrating a deep continuity from basal to abstract higher-level cognition.

PP argues that prediction-error minimization is the mechanism underlying all perception, action, and cognition. This means that a hierarchical generative model of person and environment generates predictions about expected input. This model is kept in line with the world by generating top- down predictions which anticipate sensory inputs, and by bottom-up error signals which update the model (Pezzulo et al., 2024). Precision-weighting, the ability to selectively up- or down-weight top-down or bottom-up information, is a critical mechanism. This basic framework, despite its simplicity, accounts for many psychological and cognitive phenomena parsimoniously. Yet, it struggles when confronted with higher- level cognition because it cannot account for the representation of abstract concepts.

The challenge of abstract concepts originates from the fact that cognition based purely in a hierarchical generative model is limited to perceptual or action-based representations, i.e., it is embodied. Embodied approaches argue that mental representations consist of the re-activation of modalities in order to represent concepts: The representation of a word like KICK involves the re-activation of those (brain) states which are active during the execution of the action itself. This postulate is in conflict with abstract concepts, which are defined by being intangible, meaning they cannot be perceived by the modalities (for example, JUSTICE, FEAR, or mathematical functions like addition). They are therefore unable to be represented by sensory modalities. This intuition is often used to motivate accounts of symbol-based mental representations (e.g., Mahon & Caramazza, 2008). Symbols are arbitrary, meaning they are disconnected from sensory input, and do not fall prey to this criticism. PP, with its emphasis on a hierarchical generative model of person and environment, posits an embodied format and therefore fails to account for abstract concepts. Yet the embodied theories of GC have successfully addressed this challenge.

GC's solution to this challenge involves, not an appeal to symbol-based cognition, but rather retains perception- and action-based cognition, demonstrating how this can nonetheless account for abstract concepts. This account involves exaptation, a process in which traits that evolved to serve one function are later re-used for a different function. Cognitive abilities which evolved to serve the function of situated action, according to GC, are exapted for abstract concepts. GC posits that conceptual representations are not amodal symbols but reenactments of sensorimotor states, selectively reactivated in service of meaning (Barsalou, 2008). This is highly reminiscent of the notion of detached models in PP. Yet, the commensurability goes beyond this basic similarity. Four core insights from GC can deliver an account of abstract concepts under PP.

First, hierarchy, which enables abstraction (one part of the challenge of abstract concepts). Abstraction involves disambiguating sensory representations of instances of a category to generalize. For example, the concept of a CUP can be represented despite sensory variability among individual cups. GC theories have addressed this by positing hierarchical descriptions like convergence zones, multi- layered, multimodal networks with increasingly general representations emerging at increasingly higher nodes (Fernandino et al., 2016; Simmons & Barsalou, 2003). PP's architecture has a similar organization in its hierarchical generative model, where precision- weighting yields category- general representations of concepts (Michel, 2022). Therefore, both frameworks account for abstraction by postulating a hierarchical organization. Yet this enables only generalizing across specific instances of a category (i.e., abstraction), not creating novel concepts without tangible sensory referents (i.e., abstractness). The solution of hierarchy therefore only addresses the first part of the challenge of abstract concepts, and abstractness requires a different solution (Barsalou, 2003; Dove, 2016).

Second, language. The challenge of abstract concepts often motivates symbol-based theories because these offer an arbitrary representational format. Yet, words function as an arbitrary, yet nonetheless embodied (e.g., inner speech) symbol that "glues" together disparate sensory fragments. Therefore, words present a representation that can gather sensory representations of dissimilar referents (Lupyan & Bergen, 2016). In PP terms, words are labels that act as high- level context cues that selectively amplify precision at particular root nodes (Lupyan & Clark, 2015). Another function of words is that they act as social tools, extending our action capabilities beyond the body, enabling distal manipulation of others' mental states (Borghi & Binkofski, 2014). Therefore, GC theories propose language as a label and as a social tool, both of which can ground abstract concepts.

Third, metaphoric mapping. One of GC's central insights has been that concrete sensorimotor representations often stand in for abstract concepts (Lakoff & Johnson, 1999). Consider spatial metaphors for time ("moving through time") or embodied mappings of valence to left (bad) and right (good) space (Casasanto, 2009). Under PP, precision-weighting can isolate the metaphorically applied aspects of a representation by, for instance, activating interoceptive reward signals while suppressing tactile and visual predictions when "grasping an idea" (Michel, 2023; Pezzulo, 2011). Metaphoric mapping therefore argues that abstract concepts

are represented by re-using concrete, tangible concepts, and this is implemented in the hierarchical generative model by precision-weighting.

Fourth, conceptual spaces. This work posits that the same neural mechanisms enabling spatial navigation are reused to represent concepts (Bottini & Doeller, 2020; Gärdenfors, 2014). Dimensions such as hue, brightness, or saturation, for example, serve as quality dimensions of a color space. Semantic similarity corresponds here to spatial proximity, and the meaning of a concept consists of its location in conceptual space (Bellmund et al., 2018). Furthermore, it has been argued that reasoning involves simulating actions and forces in these spaces. This is evidenced by semantic foraging behaviors (e.g., random word generation tasks) that mirror animal foraging strategies (Hills et al., 2015; Malaie et al., 2023).

Together, these insights from PP and GC propose perception- action simulations to underlie even abstract concepts. This synthesis produces an empirically supported, interdisciplinary, and parsimonious extension of the current PP literature. It explicates the representational content of PP's hierarchical generative model by drawing on insights gained from GC's long theoretical and empirical tradition. Furthermore, by embedding this integration into the biogenic approach we can trace a clear continuity between the principles guiding the emergence of life and mind, and the sophisticated higher-level cognition of humans. This has consequences beyond PP because it contributes to understanding the origin of intelligent behavior generally, irrespective of organism. We argue for a phylogenetic progression of organisms: being a model, having a model, detaching models, and exapting models.

Ayoob Shahmoradi

15.3 - REPRESENTATION, VERIDICALITY, AND ACTION

Ayoob Shahmoradi, Ruhr-Universität Bochum

Abstract:
Some philosophers hold that perception is not bound by veridicality constraints—that is, perceiving an object does not require representing any of the properties the object actually has. I argue (i) that the standard positive arguments for this view are not entirely successful, and (ii) that there are prima facie plausible reasons to think some veridicality constraint on perception is needed to explain both its action-guiding role and its representational character. Along the way, I also propose a pluralistic account of perceptual reference—one that leaves open the possibility of other forms of successful reference beyond those constrained by veridicality—and I elucidate the relations between these different types of perceptual reference.