
European
Society for
Philosophy and
Psychology


Filosofický ústav AV ČR
Institute of Philosophy

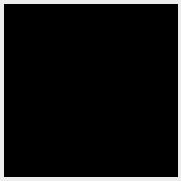
30th CONFERENCE OF THE EUROPEAN SOCIETY FOR PHILOSOPHY AND PSYCHOLOGY

PRAGUE 28/08–31/08 2023


E S P P
30th Conference of the European Society
for Philosophy and Psychology
2 0 2 3

CONTENTS

KEYNOTE ADDRESSES AND PLENARY SYMPOSIA



KEYNOTE ADDRESS #1

Teresa McCormack (Queen's University Belfast)

Can psychologists benefit from talking to philosophers about time?

The ESPP is an interdisciplinary conference premised on the assumption that it can be mutually beneficial to talk to academics from other disciplines. In this talk, I use time as a case study for considering the benefits for psychologists of interdisciplinary research with philosophers. A central debate in the philosophy of time is that between so-called A-theorists and B-theorists. This is primarily a metaphysical debate about the nature of time, but within the debate there are frequently-made claims concerning 'everyday' ways of thinking about and experiencing time that, on the face of it, look to be relevant to psychologists. I will consider three recent strands of empirical research that I have conducted in collaboration with philosophers that have all been sparked by this philosophical debate: research on (i) relief; (ii) asymmetries in attitudes to past versus future events; and (iii) people's intuitive theories of time. I will discuss the findings of each of the strands, and consider whether or not it has been beneficial to have this research informed specifically by philosophy and if so why.

PLENARY SYMPOSIUM

Lia Kvavilashvili (University of Hertfordshire):

Spontaneous future thinking and its functions in everyday life

Our remarkable ability to mentally envisage the future has been studied primarily with a simple laboratory method of asking participants to think of a plausible, but novel future event in response to random cue words. The focus of this talk will be on future thinking in everyday life to answer the question about why do people think about the future outside the laboratory, what are the typical and most frequent contents of such future thoughts, how do they occur and do they help us get things done or accomplish our goals? Relevant empirical studies across several domains (e.g., episodic future thinking, mind-wandering, prospective memory) will be selectively reviewed and a pragmatic dual process approach to future thinking will be outlined.

Cristina Atance (University of Ottawa):

Heading home: Characteristics and functions of future thinking in children's daily lives

Children's mental time travel into the future has mostly been measured using standardized, structured laboratory tasks. Although such tasks have led to important discoveries, including when specific future-oriented capacities emerge, they cannot tell us about the functions future thinking serve in children's daily lives. Gaining traction on this issue requires a different kind of approach: a naturalistic one that captures children's "lived experience." I describe a recent study from our lab in which parents of 3- to 7-year-olds tracked their children's future statements and actions during a one-week period. Using thematic analysis, we identified four functions of mental time travel into the future in children's daily lives, including (1) information seeking, (2) predicting, (3) expressing desire/intent, and (4) problem solving. I conclude my talk by comparing and contrasting these functions with those described in adults.

Christoph Hoerl (University of Warwick):

Memory and anticipation

Rather than seeing them as involving two quite different types of mental state, philosophers as well as psychologists have recently argued that episodic memory and episodic future thinking should be seen as manifestations of a unitary capacity for 'mental time travel'. In this context – sometimes using the slogan 'memory is for the future' – it has also been claimed that the function of episodic memory is to be explained in terms of its capacity to provide information useful for guiding future action. After offering some critical

observations regarding existing articulations of these ideas, I suggest a more complex picture, which also recognizes a crucial motivational dimension attached to episodic memory and episodic future thought. I argue that this is bound up with more fundamental asymmetries in the role the past and the future play in people's lives than are typically recognized in the literature on 'mental time travel'.

KEYNOTE ADDRESS #2

Nick Shea (University of London)

Concepts at the interface

Research on concepts has concentrated on the way people apply concepts online, when presented with a stimulus. Just as important, however, is the use of concepts offline, when planning what to do or thinking about what is the case. For example, when an agent thinks about making a sandwich and plans to use a knife, they already activate motor programs appropriate to grasping the knife, before they see it. There is strong evidence that inferences driven by conceptual thought draw heavily on special-purpose resources, both for concrete objects (e.g. sensorimotor affordances) and abstract categories (e.g. affective responses and valence). At the same time, concepts afford general-purpose recombination and support domain-general reasoning processes – phenomena that have long been the focus of philosophers. There is a growing consensus that a theory of concepts must encompass both kinds of process.

Because concepts act as an interface between domain-general reasoning and special-purpose systems, concept-driven thinking is able to take advantage of the complementary costs and benefits of each. This paper argues that reliance on special-purpose systems allows cognition to avoid the infamous frame problem and to retrieve information in a relevance-based way. At the same time, performing domain-general computations on conceptual thoughts enables our thinking to transcend the limitations of the if-then dispositions trained into special-purpose systems as a result of experience (Gallistel & King 2010). The ability to move flexibly between these two styles of computation makes human cognition especially powerful.

PLENARY SYMPOSIUM

Edouard Machery (University of Pittsburgh):

By default

In this talk, I will examine whether concept retrieval is fully contextual or whether concepts have a core that is, in a sense to be specified, accessed by default (a view I have called "invariantism"). Evidence from behavioral and neuroimaging studies will be used to support invariantism.

Brad Mahon (Carnegie Mellon University):

Rethinking domain-specificity in mind and brain

I discuss visual object recognition through the lens of how "downstream" systems interact with the outputs of visual recognition processes. Those downstream processes include conceptual interpretation, grasping and object use, navigating and orienting in an environment, physical reasoning about the world, and inferring future actions and the inner mental states of agents. I will emphasize functional neuroimaging studies in patients with acquired brain lesions to show that lesions outside the ventral visual object processing stream can cause category-specific patterns of information disruption within the ventral stream. I will argue that those findings support the view that connectivity between ventral stream object recognition areas and other regions of the brain is the basis for the emergence of neural specificity for a limited number of semantic domains in the brain.

Gabriella Vigliocco (University College London):

Situating conceptual processing in the real world

The ecology of conceptual processing is, for the most, multimodal: we talk about things while we are looking/doing them; we talk about things to others. In the talk, I will present results from recent work in which we ask: (1) is the processing of concrete and abstract concepts the same when we process them in a situated or displaced context? (2) how does social interaction change the way in which we process concepts?

KEYNOTE ADDRESS #3

Peter Hagoort (MPI Nijmegen)

The natural kind problem in cognitive neuroscience

In cognitive neuroscience we try to relate models of the cognitive architectures of key functions, such as memory and language, to models of the neural architecture of the human brain. To arrive at the relevant level of understanding in cognitive neuroscience, this requires answers to the following questions:

(i) What are the relevant natural kinds for cognition (NKC)?
(ii) What are the relevant natural kinds at the level of the brain (NKB)?
(iii) What is the right mapping relation between NKC and NKB?

I will discuss these questions in more detail for the relation between brain and language.

I will argue that taking a stance on the relevant natural kind issues is crucial for progress in cognitive neuroscience.

PLENARY SYMPOSIUM

Raffaella Bernardi (University of Trento):

The interplay between language generation and reasoning in Large Language Models: Information seeking game as test-bed

Large Language Models (LLMs) have recently grabbed the attention of the community and the media. Having reached high language proficiency, attention has been shifting toward its reasoning capabilities. In this talk, we employ the 20-Questions game, traditionally used within the Cognitive Science community, to inspect the information seeking-strategy's of LLM. This task requires a series of interconnected skills: asking informative questions, stepwise updating the hypothesis space, and stopping asking questions when enough information has been collected. We show where LLM stand with respect of each of them. Moreover, we consider both the traditional language based version of the game and its visually grounded counterpart discussing future challenges.

Aida Nematzadeh (Google Deepmind):

On evaluating neural representations

There has been an increased interest in developing general-purpose pretrained models across different domains, such as language, vision, and multimodal. This approach is appealing because we can pretrain models on large datasets once, and then adapt them to various tasks using a smaller supervised dataset. Moreover, these models achieve impressive results on a range of benchmarks, often performing better than task-specific models. Finally, this pretraining approach processes the data passively and does not rely on actively interacting with humans.

In this talk, I will first discuss what aspects of language children can learn passively and to what extent interacting with others might require developing theory of mind. Next, I discuss the need for better evaluation pipelines to better understand the shortcomings and strengths of pretrained models. In particular, I will focus on commonsense reasoning, verb understanding, and theory of mind as challenging domains for our existing pretrained models.

Mariya Toneva (Max Planck Institute for Software Systems):

Language modeling beyond language modeling

Language models that have been trained to predict the next word over billions of text documents have been shown to also significantly predict brain recordings of people comprehending language. Understanding the reasons behind the observed similarities between language in machines and language in the brain can lead to more insight into both systems. Additionally, the human language system integrates information from multiple sensory modalities which puts text-only language models at a fundamental disadvantage as cognitive models.

In this talk, we will discuss a series of recent works that make progress towards these questions along different dimensions. The unifying principle among these works that allows us to make scientific claims about why one black box (language model) aligns with another black box (the human brain) is our ability to make specific perturbations in the language model and observe their effect on the alignment with the brain. Building on this approach, we find that the brain alignment can be improved by training a language model to summarize narratives, and by incorporating auditory and visual information from an ongoing event. Taken together, these works make progress towards determining the sufficient and necessary conditions under which language in machines aligns with language in the brain.

KEYNOTE ADDRESS #4

Ophelia Deroy (LMU Munich)

Shared experiences and perceptual common ground

You and I are looking at the Vltava from a panoramic terrace. We see Charles Bridge and its statues, the tower, Mánes Bridge, the castle and roofs in the background. It seems to make sense to say that we had a shared experience of the scenery. Or does it? In this talk, I present a new theoretical and empirical framework for shared experiences, which shows – at least – three things:

(1) that sharing perception is not the same as jointly attending

(2) that the effects of shared experiences, sometimes documented as "amplification", are not just explained by synchrony

(3) that sharing perception is not just a matter of adopting an attitude towards our own experience, or placing its contents in common ground.

If you want to know the positive definition I propose, come and listen to the talk.

PLENARY SYMPOSIUM

Bahador Bahrami (LMU Munich):

Shared responsibility in collective decisions

Research investigating collective decision-making has focused primarily on the improvement of accuracy in collective decisions and less on the motives that drive individuals to make these decisions. We argue that a strong but neglected motive for making collective decisions is minimizing the material and psychological burden of an individual's responsibility. Making difficult decisions with others shields individuals from the consequences of negative outcomes by reducing regret, punishment and stress. Considering shared responsibility as a key motivation to join groups helps understand behaviours with societal implications such as political voting, committing norm violations, predicting natural disasters and making health-related decisions.

References

1. El Zein, M., & Bahrami, B. (2020). Joining a group diverts regret and responsibility away from the individual. Proceedings of the Royal Society B: Biological Sciences, 287(1922), 20192251.

2. El Zein M, Dolan RJ, Bahrami B. Shared Responsibility Decreases the Sense of Agency in the Human Brain. J Cogn Neurosci 2022; 34 (11): 2065–2081.

3. El Zein, M., Bahrami, B., & Hertwig, R. (2019). Shared responsibility in collective decisions. Nature Human Behaviour, 3(6), 554–559.

Olle Blomberg (University of Gothenburg) and Björn Petersson (Lund University):

Collective faults and guilt feelings in unstructured groups

On a number of accounts of collective moral obligation, an unstructured group can have a moral obligation that is not reducible to a set of individual moral obligations (see e.g. Blomberg & Petersson 2023). When the group members fail to live up to such a collective moral obligation, it is possible on these accounts that one or more members may not individually be at fault or have a substandard "quality of will" even though the group is blameworthy—they may have done everything in their power to get the rest of the group to do their parts of what is needed for the group to act in accordance with the obligation. In this talk, we sketch an account of collective blameworthiness, and address the following questions: How should personally faultless group members appropriately respond as members of a collectively blameworthy group? And what is the appropriate stance for blamers to take toward such individually faultless group members? Drawing on work in social psychology on group identification and collective guilt (e.g. Hogg, Abrahams & Brewer 2017; Branscombe & Doosje 2004) as well as our own account of collective moral obligation, we argue that it is at least rationally permissible for group members, including those not individually at fault, to feel guilt from the group's perspective, in light of the group's failure to act in accordance with its obligation. Similarly, it is rationally permissible to blame the group members, qua

group members. We compare and contrast this view to views according to which collective guilt is fitting because guilt simply does not imply perceived fault or blameworthiness (Morris 1987; Sepinwall 2011; Woo 2023), or according to which some other moral emotion than guilt is fitting, at least when it comes to individuals who are not at fault or whose quality of will is not substandard (e.g. Oshana 2006; Björnsson 2021; Telech 2022), as well as accounts claiming that assignments of collective guilt have no implications for individual members' guilt (Gilbert 2000; Cooper 2001).

References

Blomberg, Olle and Björn Petersson. 2023. 'Team Reasoning and Collective Moral Obligation', *Social Theory and Practice*. Advance online publication, January 27. https://doi.org/10.5840/soctheorpract2023120177

Björnsson, Gunnar. 2021. 'Being Implicated: On the Fittingness of Guilt and Indignation over Outcomes'. *Philosophical Studies* 178 (11): 3543–60. https://doi.org/10.1007/s11098-021-01613-4.

Branscombe, Nyla R and Bertjan Doosje (eds). 2004. *Collective Guilt: International Perspectives*. Cambridge: Cambridge University Press.

Cooper, David. 2001. 'Collective Responsibility, "Moral Luck," and Reconciliation' in *War Crimes and Collective Wrongdoing: A Reader,* ed. Aleksandar Jokic, Malden: Blackwell.

Gilbert, Margaret. 2000. *Sociality and Responsibility: New Essays in Plural Subject Theory*. Lanham: Rowman & Littlefield.

Hogg, Michael A, Dominic Abrams, and Marilynn B Brewer. 2017. 'Social Identity: The Role of Self in Group Processes and Intergroup Relations'. *Group Processes & Integroup Relations* 20 (5): 570–81. https://doi.org/10.1177/1368430217690909.

Morris, Herbert. 1987. Nonmoral Guilt. In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology* (pp. 220-240). New York: Cambridge University Press.

Oshana, Marina A. L. 2006. 'Moral Taint'. *Metaphilosophy* 37 (3–4): 353–75. https://doi.org/10.1111/j.1467-9973.2006.00437.x.

Sepinwall, Amy. 2011. 'Citizen Responsibility and the Reactive Attitudes: Blaming Americans for War Crimes in Iraq'. In *Accountability for Collective Wrongdoing*, eds. by Tracy Isaacs and Richard Vernon, 231–60. Cambridge University Press. https://doi.org/10.1017/CBO9780511976780.010.

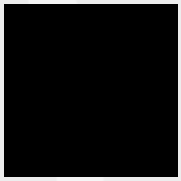
Telech, Daniel. 2022. 'Relation-Regret and Associative Luck' in Morality and Agency: Themes from Bernard Williams, eds. A. Szigeti and M. Talbert, Oxford University Press.

Woo, Jaeha. 2023. 'Moral Guilt without Blameworthiness'. *Southwest Philosophy Review* 39 (1): 201–8. https://doi.org/10.5840/swphilreview202339122.

Elisabeth Pacherie:
Variety and unity of shared experiences

We are conscious creatures but we are also hyper-social creatures. Indeed, some of our more vivid experiences appear to be shared experiences: the joy experienced by supporters when their team wins a game, the sense of joint agency experienced by the team members when playing together or the memories evoked when old war veterans reminisce together. However, the exact nature of such shared experiences remains difficult to pin down. In what sense must experiences be related to qualify as shared experiences? What specific subjective qualities, if any, characterize them in contrast to purely individual experiences? Is there a phenomenology of we-ness common to different kinds of shared experiences and different kinds of we? What are the cognitive processes responsible for their emergence? What are the factors modulating these experiences? How can they be studied experimentally? I will try and provide some tentative and partial answers to these questions and highlight some of the many outstanding questions we face.

CONTRIBUTED SYMPOSIA



The many forms of colour vision

This symposium brings together three researchers who all work in the interdisciplinary field of colour research, here representing the fields of colour psychophysics, neurophysiology, colour development, and philosophy of perception. Although their individual topics may seem diverse, the overarching (or perhaps background) topic concerns the diversity of colour vision—of what it takes to see colour in a highly complex material world (consisting of coloured surfaces, illuminants and volumes), of the kinds of colour experience there are and what use they might be, and how such a complex set of visual capacities comes to comprise 'the' human capacity for colour vision.

Robert Kentridge, a colour psychophysicist and expert on cerebral achromatopsia, will speak to the issue of colour experience. As a visual capacity, colour vision allows us to find, track and identify objects and their materials. But qua a form of visual experience, its utility is less discussed. Why exactly do we experience the reds, greens and blues? Will Davies will discuss modes of colour diversity. There are many things to which we assign colour, some with more confidence than others: surfaces of objects (for a start, think of the 5–7 types of indoor paint you are typically offered in home improvement stores— matte/flat, gloss, semi-gloss, eggshell or satin; now add the multitude of surface textures in nature), volumes (ice cubes, a bath tub of water, translucent jellyfish, milky coffee, fog, oceans) and illuminants both natural and artificial (sunlight, moonlight, fire, bio-fluorescence). All these factors contribute to colour appearance. They also, each and every one, constitute different forms of colour perception. Or do they? Here, Prof Davies takes up this question and suggests a several different ways to parse this territory. The third and final talk, Akins and Hahn discuss colour development, in effect what it would it mean to bring a infant, then toddler, then small child up to snuff in seeing a world of colour. Because most people think of vision in terms of a exclusive dichotomy— 'either in black and white or colour'—colour development poses both a philosophical and an empirical puzzle. If infants come into the world seeing in black and white, when and how do does colour vision arise? Presumably, not suddenly one day (overnight?). But if not, how would colour develop incrementally? By breaking down colour development into the visual capacities required to see in colour, this talk seeks to show how colour vision can development 'bit by bit'. And why that makes sense.

Will Davies (University of Oxford)

The Heterogeneity of Colour Perception

Philosophical orthodoxy treats colour perception as autonomous and homogeneous.

Autonomous, in that colour perceptions are explained independently of object perceptions.

Homogeneous, in that colours appear in the same fundamental ways in all perceptions, regardless of the type of object one perceives. Gestalt psychology, in contrast, considered colour perception nonautonomous, positing explanatory interdependencies between perceptions of colour, objects, and space. Accounts of perceived colour proceeded in lockstep with the perceptual articulation and organisation of the scene into different types of object, including material surfaces, transparent volumes, illuminants, and illuminations. Colour perception was deemed heterogeneous, in that different types of perceived object have different ways of appearing coloured, or different modes of colour appearance, in David Katz's (1911/1935) phrase. These included the film or aperture mode wherein colours appear as features of indefinitely located expanses with no apparent materiality. The Optical Society of America (1953) recognised five modes, enshrining their importance to colour theory. Seventy years on, however, there is little understanding of how (if at all) they bear on philosophical issues concerning colour.

I shall consider how best to articulate and explain these putative differences among modes of colour appearance. The Gestalt psychologists emphasised phenomenological differences, making many subtle observations that merit closer attention. I focus in particular on the different ways that colours can appear to occupy regions of space, depending on the type of object to which the colour is attributed. This develops the close connection between colour perception and spatial representation. Modern vision science provides further clues. Optics helps to clarify differences among the physical magnitudes causally

implicated in perceptions of the colours of surfaces, transparencies, and illuminations. For instance, opaque surfaces and transparencies affect the spectral and intensive properties of incident light in different ways, via their specular and diffuse reflectance and transmittance characteristics. Illuminations comprise light fields wherein light of varying spectral and intensive qualities arrives at each point in the scene from every direction.

Psychophysics investigates our capacities to discriminate and scale these magnitudes, providing mappings from magnitudes to perceived colours, along with other closely related features like glossiness, translucency, and light flow. It also produces representations of the colour spaces associated with stimuli in each domain, such as spectral lights, material surfaces, or (in principle) transparent volumes. Drawing on these diverse resources, I sketch several ways one might individuate modes of colour appearance. The resulting views commit one to either pluralism about colour perception or about colour, perhaps both.

Robert Kentridge (Durham University)

What Use is Colour Experience?

Colour is a confusing topic for many reasons. One is that we use the term both to describe an aspect of our visual experience and to describe a property of the surface of objects in the world which is related to the way in which they reflect light of different wavelengths. These usages are often in conflict. We may judge that two objects are covered in the same colour of paint but, if one is in sunlight and the other in shadow, the colour experiences they elicit will differ. As the function of vision is most obviously to inform us about the properties of objects in the world (such as the nature of their surface material) so that we can act in a manner that maximises our chances of survival, we might ask what the function of that other version of 'colour', colour experience, could be? Our colour experience neither matches the proximal colour stimulus (the light reaching our eyes) nor the distal stimulus (the spectral reflectance of an object's surface).

By analysing how and when colour experience is likely generated, and in addition considering the timing of experience in general, I suggest a role for colour experience (and phenomenal experience more broadly) in episodic memory.

Kathleen Akins & Martin Hahn (Simon Fraser University)

Colour Bit-by-Bit: On the Development of Colour Vision

One of the oddest facts about human colour vision is just how long it takes for children to begin using colour terms correctly. In Darwin's time, children would typically 'know their colours' between the ages of 5-7. Contemporary children have colour terms in their vocabulary from about age two years onwards, but do not use those terms reliably until about 3-5 years of age, a full two years younger. Whether now or then, children do not learn colour words in one leap. Usually a colour term will be 'applied' to a favourite toy or plush/stuffed animal; hence the words 'blue' applied to Thomas the Tank Engine or 'pink' applied to plush pink rabbit. Thereafter, there is no sudden uptake in colour terms; colour names are learned one by one over a long and seemingly arduous process. (Darwin himself, recorded his worries about his children, whether or not they were colour blind. He wrote (Darwin 1877): "….soon after they (Darwin's children) had come to the age when they knew the names of all common objects, I was startled by observing that they seemed quite incapable of affixing the right names to the colours in coloured engravings, although I tried repeatedly to teach them."). But stranger still, prior to learning colour terms, young children can use colour. They can match colours to samples ("Do you have one like this?"), to make inferences about object categories based on colour ("What do you think this thing does?"), and even sort cards by colour category ("In which pile do

you think this one should go?") (Miller & Johnston- Laird, 1976). They can also pass a children's version of the Ishihara colour vision test, recognizing coloured figures disguised by luminance noise.

The topic of this paper will not be about colourlanguage, which merely serves to highlight the puzzle, but about colour development. We know that neonates do not have the capacity for colour vision. As the infant grows into a toddler and young child, somehow colour vision is acquired. But what does it mean to acquire colour vision? Do children wake one morning to find a coloured world? Probably not. But what could it mean to learn to see the colours bit-by-bit, a conclusion that must surely be true, given everything we know about developmental neuroscience. Apart from the control of respiration, neural capacities do not suddenly switch on, after a protracted period of neural development. Rather, they begin in a nascent form, sometimes then switch forms, and then gradually develop into the adult capacity. But how is this even possible? This is a philosophical question above all else. If we think that vision must be in 'in black and white' or (exclusive) "in colour", it seems impossible to resolve this dilemma. But it is also a physiological conundrum. By looking at the many capacities required for colour vision and their development, we resolve both these puzzles.

Urte Laukaityte (UC Berkeley), David Harrison (University of Cambridge), and Matthew Sims (RUB)

The Biogenic Account: New perspectives in understanding the material basis of cognition

One can learn much from a cursory glance at the history of cognitive science. Nominally composed of six interrelated disciplines, it is well remarked that certain subdisciplines—notably psychology, linguistics and artificial intelligence, and later neuroscience—have had a disproportionate effect on the development and trajectory of the field as a whole. Thus, from the so-called 'cognitive turn' catalysed by Miller's cognitive revolution (Miller, 2003) and Chomsky's 'linguistic turn', through to the development of cognitive neuroscience as a standalone discipline, certain areas of research have been privileged in their influence and dominance in the field of cognitive science. Nevertheless, there appears to be a unifying thread connecting these distinct iterations of cognitive science, and that is the relative neglect of the materiality and material embeddedness of the mind-brain. In our symposium, we would like to suggest that the materiality of the body, brain, mind, and environment has not been taken far enough. That is, we articulate and discuss how and why the emerging fields of basal cognition, active matter physics, and soft robotics should have relevance in sculpting, shaping, and sketching the future of cognitive science.

Briefly, basal cognition is a new area of research that assesses the cognitive and computational capacities of 'basal' organisms: where this is typically taken to mean organisms that lack nervous systems or where nervous systems are present in only a minimal sense (e.g., in comb jellies). Standard model organisms consist of acellular or cellular slime moulds, colonial bacterial organisations, basal animals such as Cnidaria or Placozoa, and, in some cases, artificially engineered, 'soft' robotic systems capable of minimal forms of cognitive capacities. The idea here is to understand how cognition, mind, and intelligence manifest and relate to the metabolic, homeostatic, and physiological imperatives required for thriving in a precarious environment. Crucially, the absence of neurons in the case of Placozoa, as well as in the case of simple neural organisms as Hydra and other Cnidaria, the absence of a centralised nervous system is telling (Ginsburg & Jablonka, 2021). This is because the goal of basal cognition research is to decentralise the importance of neurons and nervous systems in manifesting cognitive capacities such as learning (Dexter et. al., 2019), goal-directedness (Levin, 2019; 2020), and prototypical instances of 'minimal' subjectivity (Godfrey-Smith, 2016a; 2016b). Indeed, as one prominent researcher in the area remarks, "capacities usually assigned to [organisms] with nervous systems, such as integrating spatio-temporal information, memory, and ability to pursue specific outcomes via selection from a number of possible behaviours evolved from far older pre-neural origins" (Levin, 2019: 2, emphasis added). Moreover, renowned neuroscientist Samuel Gershman (Gershman et. al., 2021: 1) has reconsidered the evidence for complex learning in unicellular organisms, suggesting that it has "the potential to profoundly reshape our understanding of learning in multicellular organisms".

One perspective comes from psychiatric disease, which offers a case study on basal cognition. Bipolar disorder, in particular, is a portrait of metabolism's impact on thought: a hereditary disorder of neuronal excitability, manifest in changes in drive and appetite— for sleep, food, sex, novelty— as well as mood shifts, racing thoughts and cognitive changes. Molecular findings about the cells of bipolar patients offer clues to what metabolic changes may modulate dimensions of thought (Mishra et al., 2021; Shen et al., 2020; Hoffman, et al, 2018; Mertens et al., 2015; O'Shea et al., 2015; etc.). One of the proposed presentations will thus assess the extent to which the biogenic account is not restricted to organisms of lower complexity (e.g., slime moulds, comb jellies, and so on) but rather can be scaled up to accommodate and elucidate features of higher mammalian cognition. Specifically, this talk focuses on the field of Affective Neuroscience, which increasingly emphasises the physiological and somatic basis of cognitive processes, and how this can fit hand-and-glove with the more theoretical work in the basal cognition literature.

Intriguingly, basal cognition has already influenced the field of soft robotics: a subdiscipline of robotics and artificial intelligence that places a premium on 'soft', 'elastic', and 'vulnerable' dimensions of embodiment, motility, and mentation (Bongard & Levin, 2021; Man & Damasio, 2019). Because researchers in the field of basal cognition associate mind with the physiological demands for survival, there also tends to be an emphasis on the 'precarious' nature of our biological embodiment (Froese 2016). Picking up on a similar theme, Man and Damasio (2019) have recently articulated a novel position vis-à-vis more traditional artificial intelligence and robotics approaches to Artificial General Intelligence (AGI). Thus, they write, "We propose the design and construction of a new class of machines organised according to the principles of life regulation, or homeostasis. These machines must have physical constructions—bodies—that must be maintained within a narrow range of viability states and thus share some essential traits with all living systems" (2019: 446). In addition to the implications basal cognition might have for the mind and brain sciences, it appears that it could motivate—via the related discipline of soft robotics—alternative approaches to more 'traditional' paths towards AGI. Indeed, this is most evident in the study of 'embodied computation', which lessens the cognitive load required of centralised processors by outsourcing to the materiality of the body itself. Furthermore, the turn away from standard "hard" materials used in robotics towards adaptive and responsive materials connects soft robotics with the growing body of research within active matter physics. This subdiscipline of physics investigates both biological and non-biological materials which are endogenously out-of-equilibrium and sustain their existence through the energy consumption of constituent elements. It is proposed, then, that the emerging field of basal cognition, together with soft robotics and active matter physics, has ramifications and significance for multiple disciplines of cognitive science.

Thus, our symposium proposes to chart, map, and discourse this exciting body of literature and assess how (indeed, whether) basal cognition research should permeate, influence, and affect the mind and brain sciences: from neuroscience to artificial intelligence. The underlying theme will centre around explicating basal cognition, its distinctive claims, potential problems, and the significance it may hold for other disciplines.

Eugen Fischer (University of East Anglia), Richard Breheny (University College London), Dimitra Lazaridou-Chatzigoga (University of East Anglia), and Kevin Reuter (University of Zurich)

The Varieties of Salience: Philosophical and Psycholinguistic Relevance

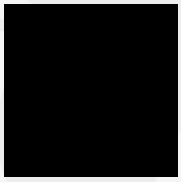
Recent studies suggest that salience is a key factor influencing judgment and reasoning in philosophically relevant contexts. If so, this philosophically neglected cognitive property requires examination for purposes ranging from conceptual analysis and conceptual engineering to the analysis of philosophical arguments. Salience is a property of word senses and semantic features (components of concepts). While the notion lacks a generally accepted definition, it is usually taken to capture a signal-to-noise ratio, such as the prominence of a feature for category C, or its availability in response to names or instances of C. To illustrate, the concept of shark may be characterized by a set of features, including <is a fish> and <attacks humans>. <fish> is necessary; in contrast, <attacks humans> is rarely instantiated but highly salient. Salience is

often investigated with sentence completion tasks and semantic feature production tasks (e.g., Hampton, 1995; McRae et al. 2005) that ask participants to name features that come to mind. Whether a feature is salient may be more important for cognitive processes like memory, categorization, recognition, and even decision-making than whether it is necessary or universal (Smith et al., 1988, Solomon & Barsalou, 2001, Vigliocco et al., 2004).

This symposium examines the cognitive relevance of salience and explores philosophical applications. The first strand seeks a more precise understanding of the notion of 'salience' and of the information about concepts that can be gleaned from semantic feature production tasks. An Introduction to which all symposiasts contribute discusses related notions and operationalisations of salience, from different disciplinary perspectives. Talk 1 (Reuter) then explores the philosophical relevance of salience by presenting findings from semantic feature production tasks concerning the concepts of conspiracy theory, life, and dinosaur, and discussing the findings' relevance to conceptual analysis and conceptual engineering.

The second strand examines salience effects in polysemy and metaphor processing, and discusses their relevance to the analysis of philosophical arguments. Studies implementing the psycholinguistic cancellation paradigm with pupillometry, fixation times, and plausibility ratings have documented contextually inappropriate inferences that are supported only by the dominant sense but are made from subordinate uses (e.g., inferences supported by the visual sense of 'see' triggered by metaphorical epistemic uses, as in 'see someone's point'; Fischer & Engelhardt, 2020). Such inferences from appearance- and perception-verbs and the noun 'zombie' have been suggested to give rise to fallacies of equivocation in philosophical arguments 'from illusion', 'from hallucination', and the 'zombie argument'. Talk 2 (Fischer and Lazaridou-Chatzigoga) presents two new eye-tracking studies that document such inappropriate inferences where the word of interest is preceded by disambiguating context. This work assumes that polysemous words automatically activate overlapping stereotypes (Brocher et al., 2018), followed by selective suppression of contextually irrelevant information; the documented inferences are explained as due to a linguistic salience bias. There are different computational models which can capture salience effects. Talk 3 (Breheny) proposes an alternative model to account for the apparent automatic salience of stereotypical inferences, in terms of priors on sources of relevance and their effect in situations of contextual uncertainty. The talk concludes by discussing possible consequences for the philosophical projects presented in earlier talks.

CONTRIBUTED PAPERS



Allegretti, Pietro (The University of Waikato – Te Whare Wananga o Waikato)

Sense of Beauty and Aesthetic Predisposition in Evolutionary Aesthetic Theorising

In this talk I argue that evolutionary aesthetics would benefit from a refined conceptual framework. I will draw attention to a case study central to the aims of evolutionary aesthetics, that is, study of the phylogenesis of secondary sexual characters. The purpose of this talk is to reestablish Darwin's account of sense of beauty (henceforth, 'SoB') as a legitimate concept of evolutionary and philosophical research, and to draw connections between Darwin's account and the work of anthropologist Ellen Dissanayake on aesthetic predispositions. I will show how this synthesis of Darwin and Dissanayake will contribute to evolutionary debates concerning the development of aesthetic characters.

Darwin's account of SoB focuses on different objects than the main concepts of evolutionary aesthetics, like the property of beauty (or state of being beautiful), or judgments of beauty (a cognitive process). Darwin's main purpose was to provide explanations for the (minimally) sense-based perceptual preferences of various species, and his work constitutes a first phylogenetic reconstruction of the components of these preferences (Darwin 1987). I argue that Darwin's SoB is an adequate first guide for identifying which characteristics, both ornamental and behavioural, are the best candidates for evolutionary aesthetic explanation. I provide a framework for characterising the Darwinian SoB's components: first, there is a formal level, expressed in the display of recurrent transpecific geometrical patterns and rhythmical repetition of movements and sounds. Second, an affective, (proto-)emotional or 'felt' level, expressed in the association of sensations of pleasure with specific perceptive patterns.

Although Darwin provided an account fruitful for the first level, he could not achieve an exhaustive account of the second level. Rather, Darwin hoped for new physiological discoveries (Darwin 1890, p. 334). To complete his wish, I argue that Dissanayake's work fits the bill. Using recent discoveries in ethology and neuroaesthetics, Dissanayake studies the development of innate aesthetic predispositions in humans—these give rise to processes of 'artification'—and she shares with Darwin the idea that these predispositions are grounded in formal patterns, emotional reactions and affective states that are common to different species (Dissanayake 2014).

Dissanayake does not explicitly connect SoB with forms of animal bodies or behaviours, but her explanation of aesthetic predispositions ranges over the displays of animals and neurological triggers of these. She shares with Darwin the hypotheses that perceptual discrimination is associated with emotional arousal and these kinds of associations trigger the so-called instinctual emotions, that is, "the hereditary effect of archaic trains of ideas on the structure of the body, permanently linked to pain and pleasure perceptions" (Bartalesi & Portera 2015, p. 105). One implication of this, I argue, is that artification is not a process exclusive to humans; I will indicate several examples of non-human artification triggered by affective responses based on perceptual discrimination. This in turn may influence theories of human aesthetic evolution.

Finally, I shall conclude by indicating a framework for future research. This will involve identifying the phenotypes able to induce pleasure and attraction in individuals and linking them to the broader evolutionary aesthetic discussion.

Almagro, Manuel (University of Valencia) and Neftalí Villanueva (University of Granada)

Can we get polarized over what we don't understand?

Polarization of public opinion has become a widespread phenomenon in contemporary democracies, raising concerns about the proper functioning of these societies (Carothers & O'Donohue 2019; Levitsky & Ziblatt 2018; McCoy & Somer 2019). However, the complexity and abstract nature of many political issues that are at the center of these divisions raise the question of whether citizens can be polarized over matters they may not fully comprehend. Can we become polarized over issues that we do not fully understand?

Consider the following cases.

Spanish Constitution. Since 2008, public opinion in Spain regarding national territorial distribution has become increasingly polarized. Studies conducted by the Centro de Investigaciones Sociológicas (CIS) found that between 2008 and 2012, public opinion on this topic became divided. Support for a single central government, an option associated with the right-wing, rose from 8.6% in 2007 to 24.9% in 2012, while support for recognizing the right of Spanish regions to independence grew from 17.4% in 2008 to 28.2% in 2012 in Catalonia. Since then, support for these options has remained relatively stable. A 2016 survey revealed that 77.7% of survey participants believed that there was no need to reform the Constitution to give more self-government to Catalonia or to modify the Spanish territorial model. Surprisingly, this survey also revealed that nearly half of the Spanish population had never read the Spanish Constitution at all, while only 15.5% claimed to have read it. How can this be possible? How could a significant percentage of the Spanish population claim to have strong convictions with respect to a text that they actually know so little about?

UK's Immigration. Before Brexit, Ipsos, a British polling company, conducted studies which found that UK citizens believed that European Union migrants comprised 15% of the country's population –Brexit supporters believed that it was 20%, while Brexit opponents estimated it to be 10%. UK citizens also believed, on average, that one in four migrants were from the European Union (Legrain 2020, 23). Taken together, these results are striking: they imply that UK citizens believed that 60% of the UK population were migrants –Brexit supporters estimated it to be 80% and Brexit opponents 40%.

Given that the issues at the heart of political polarization are often quite complex, it's not surprising that many people aren't fully aware of the intricacies surrounding a given issue, or the potential outcomes of different policy decisions. And yet, we often express strong opinions about matters we know very little about. How is this possible? How can we be so deeply divided over issues that we don't fully comprehend?

These questions give rise to a previous one, which belongs to a classic philosophical puzzle: Can we believe what we do not understand? Our train of thought and our everyday conversations are full of concepts with respect to which our knowledge is very limited. We talk about elms and beeches, arthritis, synecdoches, GPT-4, depression, etc., and our utterances are systematically evaluated as true or false even if neither we, nor our audiences, can set elms and beeches apart, know exactly how to characterize arthritis or depression, or have the slightest idea of the algorithm behind GPT-4. How is the content of our thoughts and utterances fixed –so that it can be evaluated as true or false– if we have, at best, imperfect mastery of the concepts that we used?

In this paper, we will build upon those questions to discuss the following ones: When we discuss issues that we do not fully comprehend, do our assertions reflect our beliefs or are they merely expressions of something else –identity, virtue, etc.? Is the content of our utterances inherited from our mental states in these cases? Addressing these questions, it seems, is crucial to tackle the initial question of whether we can become polarized over issues that we do not fully understand.

We will focus on two competing hypotheses regarding whether we can believe what we don't understand. The first hypothesis maintains that our assertions about complex issues that we do not understand function as a way to display our identity (see Davies 2020; Eastman & Stein 1993; Goffman 1969; Hampel & Irions 2015; Kurzban & Christner 2011; Mercier 2020; Petersen 2020; Simler & Hanson 2017). According to this hypothesis, we can make assertions that express beliefs we may not hold just because we want to demonstrate our loyalty to certain groups. Let's call this the Identity Display Hypothesis (IDH henceforth).

The second hypothesis, stemming from a more traditional debate, holds that the content of our assertions about issues we don't fully understand is determined independently of our conceptual mastery, through semantic deference (Recanati 1997, 2000; Woodfield 2000; see also Sperber 1975, 1985; Villanueva 2005). According to this second hypothesis, we can hold and express beliefs that we do not fully understand. We shall refer to this as the Deferential Beliefs Hypothesis (DBH henceforth).

To determine which hypothesis is more plausible in explaining cases such as Spanish Constitution or UK's Immigration we need to examine in turn the strengths and weaknesses of each hypothesis, evaluate their explanatory power, and analyze the reasons to hold IDH or DBH. That's the goal of this paper. In particular, we consider the implications of these hypotheses for the current distinction between doxastic and affective polarization.

We will argue that IDH is in a better position to explain the current state of polarized democracies than DBH is. This is because IDH suggests that data from surveys designed to measure doxastic polarization, such as the examples introduced at the beginning, are actually examples of affective polarization, which

aligns better with the existing evidence regarding the present condition of many democracies worldwide. The bulk of our analysis will focus, thus, on the theoretical and practical advantages and disadvantages of both competing approaches.

Alvarez, Juan F. (Centre for Philosophy of Memory, Université Grenoble Alpes)

Remembering and relearning: A compatibilist view

At least to a first approximation, remembering involves encoding, storing, and retrieving information about personally experienced events. Relearning involves encoding, forgetting, and re-encoding information about personally experienced events. Martin and Deutscher (1966) viewed remembering and relearning as incompatible. This view—call it "incompatibilism"—is widely endorsed in philosophy of memory. It is endorsed by causalists (Bernecker 2010), simulationists (Michaelian 2016a), and functionalists (Fernández 2019) alike. In this paper, I argue for two claims. First, incompatibilism is a dogma; that is, it is a view held by philosophers without any proper argument supporting it. Second, "compatibilism", the view that remembering and relearning are in principle compatible, is a preferable option.

I argue that incompatibilism is a dogma in §1. While philosophers endorse incompatibilism, there is no proper argument for this view in the literature. Incompatibilism is rather an intuition that has motivated the construction of the causal theory of memory (CTM). Consider a prototypical relearning case adapted from Martin and Deutscher (1966: 180):

Some time ago, Kent saw a car accident. He told his friend, Gray, all about the accident. Later he forgot all about this event. Gray realized that Kent does not remember the accident and tells him about it. Then, Kent forgets being told about the accident but still remembers his friend's description. Today, Kent claims to remember the car accident.

According to advocates of CTM, cases of this kind elicit the intuition that relearning subjects do not remember because the memory of an event is necessarily caused by the experience of that event and, in relearning, this is not the case (see, e.g., Bernecker 2010; Debus 2017; Martin and Deutscher 1966). The causal-aetiological requirements of remembering are thus incompatible with relearning. Based on this intuition, causalists analyze remembering in such a way that cannot possibly co-occur with relearning. However, since it is unclear whether this intuition is universal ([reference omitted for review purposes]) and there is no other argument for the view, incompatibilism is a substantive claim in need of closer examination.

In §2, I argue for compatibilism by advancing three conditional arguments. I argue that compatibilism follows if certain empirically-oriented theories of memory are right. First, if the causal theory of constructive memory (Robins 2016; Sutton and O'Brien 2023) is right, then memory traces are distributed event features. Importantly, if traces are distributed event features, then the causal history of remembering that makes it incompatible with relearning does not always obtain. At least in some cases, remembering and relearning will instantiate the same causal-aetiological requirements. Second, if trace minimalism (Werning 2020) is right, then vicarious experiences can ground remembering. Once one accepts the possibility of vicariously grounded remembering, one must also accept that "vicarious memories" (Pillemer et al. 2015) and indeed some cases of relearning constitute instances of genuine remembering. Third, if the simulation theory of memory (Michaelian 2016b) is right, then memories can be formed on the basis of non-experiential sources, such as testimony. If non-experiential remembering is possible, the compatibility between remembering and relearning is also a serious possibility.

Of course, for those who reject these theories of memory, these three arguments may be unconvincing. To provide a more persuasive argument, I discuss empirical work on "distributed memory" (Sutton et al. 2010) and "collaborative remembering" (Meade et al. 2018) in §3. The basic idea is that this work is part of a well-established research program in psychology in which the very distinction between remembering and relearning plays no role. Moreover, this psychological work overlooks whether there is some kind of (in)compatibility between the encoding, storage, and retrieval of information, on the one hand, and the encoding, forgetting, and re-encoding of information, on the other. Despite the fact that both the

remembering-relearning distinction and incompatibility are overlooked, work on distributed memory and collaborative remembering has shed light on important mnemic phenomena, such as the role of conversational interaction in the development of episodic memory (Nelson and Fivush 2004), how older adults cope with memory impairments through socially distributed memory systems (Barnier et al. 2014), the mnemic roles of technology in amnesic patients (Loveday and Conway 2011), etc.

In §4, I sketch some consequences of compatibilism for two debates in philosophy of memory: the causalism-simulationism debate and the memory error debate. On the one hand, if compatibilism is right, then CTM rests on shaky ground as incompatibilism is the basic motivation for that theory. On the other hand, if compatibilism is right, then relearning should not be counted as a memory error and, pace Robins (2020), taxonomies of memory errors that include relearning are mistaken.

Álvarez-González, Aarón (University of Barcelona)

Bi-Level Evaluative Epistemology

During the last decades we have witnessed a debate on the epistemic nature of emotion with two clear factions: those who defend the view that emotions are epistemically akin to perception and those who deny it. Call the former, epistemic perceptualism about emotion (EPE from now on); call the latter, epistemic non-perceptualism about emotions (NEPE from now on). EPE claims that emotional experience (e.g., anger about the assassination of an innocent) epistemically relates to the justification of evaluative judgments (e.g., the judgment that this is unjust) in the same way in which perceptual experiences (e.g., seeing a red sphere-like object in your surroundings) epistemically relates to the justification of existential beliefs (e.g., the belief that there is an apple in this room).

As noted by Adam Carter (2019), EPE proposes a modest foundationalism. Foundationalism because emotions, like perceptions when it comes to factual beliefs, would be the self-justifying or not-in-need-of-justification bases whose endorsement justifies non-inferentially basic evaluative beliefs, conferring them the status of justified beliefs which, when true, yields knowledge. Modest because the foundation is not an infallible experience like the Cartesian cogito, but a defeasible and prima facie foundation. NEPE denies that emotions can play an epistemically foundational role.

As it is common in philosophical debates, this one has reached a point of stagnation and impasse with defenders of EPE and NEPE reiterating their reasons in a failed attempt to convince each other. In this paper I propose a way out of that impasse: based on Ernst Sosa's idea of animal and reflective knowledge (Sosa, 1985, 2007), I propose a similar distinction for the epistemic role of emotion. That way, we can vindicate the kernel of truth in EPE and NEPE, whose conjunction would offer an exhaustive and complete analysis of emotion's epistemic contribution.

Long story short: Animal emotions produce, when justified and fitting, animal evaluative knowledge. Reflective emotions produce, when justified and fitting, reflective evaluative knowledge. EPE provides a philosophical analysis of the former, NEPE provides a philosophical analysis of the latter. Under this description, defenders of EPE and NEPE do not disagree with each other, they are just offering an analysis of different but complementary epistemic achievements, which have different but compatible epistemic statuses and different epistemic requirements. Under this description, we can take theoretical items from both EPE and NEPE to understand emotion's contribution to human evaluative knowledge.

EPE's and NEPE's insights are not only compatible but also complementary. If we just accepted the new virtue-externalist version of EPE proposed by Carter, then our picture of the epistemic role of the emotions would be incomplete. As Sosa has noted, there is something epistemically admirable in reflection, in having a reflective epistemic perspective on your animal knowledge (1983, 2015). I claim that there is also something epistemically valuable in evaluative reflection. For this reason, emotion's epistemic contribution cannot be exhausted by some virtue reliabilism on emotion, as the one articulated by Carter (2019). We also need to account for the value of having an epistemic perspective on one's axiological situation.

Amirhaftehran, Saba (PhD student at Georg-August-University Göttingen/Georg-Elias-Müller-Institute for Psychology/cognitive developmental), Marina Proft (Georg-August-University Göttingen/Georg-Elias-Müller-Institute for Psychology/cognitive developmental psychology) and Hannes Rakoczy (Georg-August-University Göttingen/Georg-Elias-Müller-Institute for Psychology/cognitive developmental psychology)

Is knowledge evaluation always faster than belief? A test of Factive theory of mind accounts

Background

What is the relation of belief ascription and knowledge ascription? Traditionally, it has been assumed that given that knowledge amounts at least to justified true belief (plus whatever additional factors may be necessary, Gettier, 1963), ascribing knowledge presupposes and involves belief ascription. In the cognitive science of mental state ascription, accordingly, it has been assumed that ascribing beliefs is primary and prior to knowledge ascription in ontogenetic, phylogenetic, and cognitive respects.

These assumptions have recently been put into question by knowledge-first epistemologies in philosophy (e.g., Williamson, 2002) and by corresponding factive Theory of Mind accounts in cognitive science (e.g., Nagel, 2017; Nagel & Westra, 2021; Phillips et al., 2021; Phillips & Norby, 2021). Factive Theory of Mind accounts claim that ascribing knowledge and other factive mental states is primarily relative to the ascription of beliefs and other non-factive mental states in evolutionary, developmental, and processing terms (such that even in adulthood, ascribing factive mental states is the default primary process relative to ascribing non-factive mental states).

One piece of evidence along the latter lines comes from recent reaction time studies (Phillips et al., 2018). In these experiments, participants evaluated the truth or falseness of a statement regarding the protagonist's knowledge ("the protagonist knows that p") or belief ("the protagonist thinks that p") after reading a vignette. The results showed that participants were significantly faster in evaluating knowledge than belief ascriptions.

These findings may show that full-fledged knowledge ascription is independent of, faster than, and thus primary relative to belief ascription. However, an alternative interpretation is the following: There is a more basic form of ascribing factive mental states (perceptual and informational access) that is indeed primary evolutionarily, ontogenetically, and cognitively – as has long been argued in the Theory of Mind literature (e.g. Call & Tomasello, 2008; Perner, 1991). But this primary form may not amount to full-fledged knowledge ascription yet (Rakoczy & Proft, 2022). One of the essential features of propositional attitudes like knowledge is their aspectuality: Objects, events, and situations can be represented under different descriptions, and it matters crucially for interpreting other agents and their knowledge under which descriptions they represent the situation in question. Oedipus may know that Yocaste is at home. Does he thereby know that his mother (identical to Yocaste) is at home? Well, not necessarily since he may be unaware of the relevant identity.

Rationale

For present purposes, this then raises the following possibility along the lines of dual process theories of Theory of Mind (Apperly & Butterfill, 2009): Perhaps subjects can and do take shortcuts when evaluating knowledge questions as long as these do not involve considerations of aspectuality (like in the previous reaction time studies). In such cases, they revert to ascribing more basic factive mental states such as informational access. In cases of belief ascription, in contrast, that in themselves raise questions of aspectuality, subjects need to use full-fledged propositional attitude concepts. This would explain the previous reaction time differences.

If, however, knowledge ascription questions were to raise issues of aspectuality as well, the ascription of full-fledged propositional attitude concepts like knowledge proper would be required. As a consequence, no reaction time differences between such modified tasks and corresponding belief ascription tasks would be expected.

Competing predictions from the different accounts would be the following: The straight factive Theory of Mind account would predict that knowledge ascription is faster than belief ascription in all cases. The

dual process account, in contrast, would predict such a pattern in non-aspectual cases only. In the aspectual cases in which full-fledged aspectual propositional attitude ascription is required in both knowledge and belief questions, there should be no such reaction time differences.

Method

To test this possibility that the reaction time to evaluate knowledge and belief varies based on the type of knowledge under question we followed up on and extended previous reaction time methods. Two different types of stories were used. First, non-aspectual (NA) stories like in Phillips et al. (2018) in which the agent was aware (True belief) or was not aware (Ignorant) of an object O at location A. In the second new type of aspectual (A) stories, the agent was always aware of object O at location A but was either aware (True belief) that O was identical to X (O=X), or not (Ignorant). Therefore, the experimental design was a 2 (story-type: Aspectual—Non-Aspectual) x 2 (Info-type: True-belief—Ignorant) x 2 ascription-type (Knowledge—Belief), with the first as between-subject and the other two as within-subject factors. Each participant goes through one story-type (A or NA) but evaluates both ascription-types (knowledge and belief) in both info-types (True-info and Ignorance). Our main interest is the interaction between story-type and ascription-type. How the reaction times to evaluate knowledge and belief differs based on story-types.

Participants (native English speakers) are being recruited and paid via Prolific platform.

Preliminary results

Data collection is ongoing. Preliminary results suggest a significant interaction between story-type and ascription-type with replication of the primary findings by Phillips et al. (2018).

Reference

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? Psychological review, 116(4), 953.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. Trends in cognitive sciences, 12(5), 187-192.

Gettier, E. L. (1963). Is justified true belief knowledge? analysis, 23(6), 121–123.

Nagel, J. (2017). Factive and nonfactive mental state attribution. Mind & Language, 32(5), 525-544.

Perner, J. (1991). Understanding the representational mind. The MIT Press.

Phillips, J., & Norby, A. (2021). Factive theory of mind. Mind & Language, 36(1), 3-26.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., ... & Knobe, J. (2021). Knowledge before belief. Behavioral and Brain Sciences, 44, e140.

Phillips, J., Knobe, J., Strickland, B., Armary, P., & Cushman, F. (2018). Evidence for evaluations of knowledge prior to belief. In CogSci.

Rakoczy, H., & Proft, M. (2022). Knowledge before belief ascription? Yes and no (depending on the type of"knowledge"under consideration). Frontiers in Psychology, 5714.

Westra, E., & Nagel, J. (2021). Mindreading in conversation. Cognition, 210, 104618.

Williamson, T. (2002). Knowledge and its Limits. Oxford University Press on Demand.

Asma, Lieke (Postdoc at Munich School of Philosophy)

How to make sense of bias in behavior and decision-making

In the literature on implicit bias, one commonly shared assumption is that a particular kind of psychological state or process, an implicit attitude, is causally responsible for implicitly biased behavior. These implicit attitudes explain why, for example, sexism, racism, homophobia persistently show up in what we do, think, and feel even though we explicitly reject them and claim to be egalitarian. Accordingly, one main aim in the literature has been to identify what kind of state implicit attitudes exactly are (see, e.g., Brownstein, 2018; Greenwald & Banaji, 1995; Holroyd, 2012; Krickel, 2018; Levy, 2015; Mandelbaum, 2016; Nanay, 2021; Olson & Fazio, 2008).

One unfortunate consequence is that much less attention has been paid to the nature of the implicitly biased behavior itself. This is noteworthy, because without a good understanding of the explanandum,

the explanatory story is bound to be missing its target, at least to some extent. The underlying idea seems to be that implicit attitudes are not only causally responsible for our implicitly biased behavior, but also characterize certain behaviors and decisions as implicitly biased. The implicitly biased behavior, broadly conceived, is taken to inherit its nature from its psychological cause. We should note, however, that in reality things are the other way around: even in the implicit association test (IAT, Greenwald et al. 1998) the participant's behavior is the object of measurement, and on the basis of this behavior an implicit attitude is ascribed to the participants. In order to understand implicit bias, then, we should pay attention to the nature of the implicitly biased behavior itself.

In this paper, my focus will be on how to make sense of bias in behavior. I start by evaluating a recent proposal by De Houwer (2019), who suggests that implicitly biased behaviors are those behaviors that are automatically influenced by cues that are indicative of the social group to which the person belongs. This understanding of the nature of bias in behavior, however, is too broad: it would suggest that I am also biased if I step of the sidewalk to make room for a person in a wheelchair. Even though we could of course claim that this is a biased response, it does not do justice to how we normally use the term 'bias'. After all, my response is fully in line with the facts and rational given the situation, and does not include an inference from the information I do have to a problematic conclusion, which seems to be the key problem of implicit bias: that we assume mental disability on the basis of physical disability for example, dangerousness on the basis of skin color, or weakness on the basis of gender.

In line with this criticism, I propose to develop an account of bias in behavior based on the work of Antony (2016). Antony maintains that bias involves going beyond the facts, but points out that this is not necessarily problematic, and in fact cannot be avoided. Biases are a response to underdetermination: the facts are consistent with an infinite number of theories, because of which we have to reduce hypothesis space through non-evidential ways (Antony, 2016, p. 161). In other words: to acquire knowledge, we cannot be fully objective. We can, however, evaluate and justify our biases by assessing whether they contribute to finding the truth, i.e., whether they are vindicated and/or ecologically valid (Antony, 2016, pp. 176-177, pp. 183-185).

Similar to acquiring knowledge, bias also plays a role in deciding what the best course of action is going to be. In many cases, the evidence does not fully determine that one course of action is the best one. When we have to select a new police chief, for example, and can choose between a (male/female) candidate that scores higher on streetwiseness and a (male/female) candidate that scores higher on formal education (see Uhlmann & Cohen, 2005), we have to reason from the information we do have – their credentials and gender – to the information we need – who is the best police chief? – to make up our minds. This is not straightforward; we have to make inferences and assumptions about what the information we have says about these candidates and how they would function as police chief. And even if we would acquire more information, many decisions involve a degree of uncertainty. The goal, then, should not be to be completely objective in deciding how to act or decide, but to use those markers that are relevant for the decision at hand.

Taking this approach, two differences with bias in acquiring knowledge, which is Antony's (2016) focus, surface. Firstly, what counts as a good action depends on your aim, and whether something counts as a good bias therefore also depends on your aim. Whether a person is in a wheelchair is relevant for whether you should make room, but not for deciding who to ask for directions, for example. Secondly, even if a bias is ecologically valid, e.g., men are generally physically stronger than women, good decision-making requires investigating relevant information. We should recognize that being a man is not part of being physically strong, while being able to bench press 25 kg more than your body weight is. If we want to know who is stronger, we should investigate strength, not gender. Bad biases block further investigation: we think we know the nature of the person based on their membership of a social group, and/or we may not realize we are using this information to make up our minds. This means that regardless of whether the bias is ecologically valid, the agent's decision making could still problematic, if the information is irrelevant in relation that in light of which the agent decides. Bad biases, then, not only lead to injustice, but also to bad decisions (cf. Antony, 2016, p. 185).

References

Antony, L.M. 2016. Bias: Friend or foe? Reflections on Saulish skepticism. In *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*, ed. M. Brownstein and J. Saul, 157-190. Oxford: Oxford University Press.

Brownstein, M. 2018. *The Implicit Mind. Cognitive Architecture, the Self, and Ethics*. Oxford University Press.

De Houwer, J. 2019. Implicit bias is behavior: A functional-cognitive perspective on implicit bias. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619855638

Greenwald, A.G. and M.R. Banaji. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*. https://doi.org/10.1037/0033-295X.102.1.4
Greenwald, A.G., D.E. McGhee, D. E., and J.L. Schwartz. 1998. Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.74.6.1464
Holroyd, J. 2012. Responsibility for implicit bias. *Journal of Social Philosophy*. https://doi.org/10.1111/j.1467-9833.2012.01565.x
Krickel, B. 2018. Are the states underlying implicit biases unconscious? – A Neo-Freudian answer. *Philosophical Psychology*. https://doi.org/10.1080/09515089.2018.1470323
Levy, N. 2015. Neither fish nor fowl: Implicit attitudes as patchy endorsements. *Noûs*. https://doi.org/10.1111/nous.12074
Mandelbaum, E. 2016. Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*. https://doi.org/10.1111/nous.12089
Nanay, B. 2021. Implicit bias as mental imagery. *Journal of the American Philosophical Association*. https://doi.org/10.1017/apa.2020.29
Olson, M.A. and R.H. Fazio. 2008. Implicit and explicit measures of attitudes: The perspective of the MODE model. In *Attitudes*, eds. R.E. Petty, R.H. Fazio, and P. Brinol, 39-84. Psychology Press.
Uhlmann, E.L. and G.L. Cohen. 2005. Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*. https://doi.org/10.1111/j.0956-7976.2005.01559.x

Baker, Austin (University of Oslo)

Nonverbal Marginalization

The nonverbal social cues that accompany speech (for example, facial expressions, gestures, and eye gaze) can be as communicatively significant as the content of the speech itself. In this paper, I identify what I argue is a very common—but philosophically unexamined—phenomenon: our tendency to allocate nonverbal cues sensitive to conversational participants' levels of respective social power such that people with more power receive comparatively more positive and affirming nonverbal cues than people with less power (e.g., looking and smiling more at men and white people in meetings). I call this 'nonverbal marginalization'. I situate the philosophical discussion in the context of the empirical literature on nonverbal cues. Often cited examples of "positive and affirming" nonverbal cues include smiling, nodding, using affirming gestures, and assuming an open and welcoming body posture while negative nonverbal cues include frowning, brow furrowing, and adopting closed body postures (c.f., Knapp, Hall, & Horgan 2013).

In section one, I sketch out the cognitive architecture of nonverbal marginalization, arguing that discriminatory patterns of nonverbal behavior can reflect and reinforce implicit and explicit biases. I argue that nonverbal marginalization caused by implicit bias tends to involve unequal distribution of positive and affirming nonverbal cues (e.g., looking and smiling more at high power people at the expense of low power people), while nonverbal marginalization caused by explicit bias tends to involve negative nonverbal cues (e.g., frowning or scowling at low power people). However, I argue that these discriminatory dynamics often occur absent explicit awareness, since the default production and interpretation of nonverbal cues occur automatically. Building on this, I also discuss joint work with [psychologist name redacted] and [computer scientist name redacted], wherein we're developing a machine learning algorithm which can track nonverbal cues in videos of naturalistic conversations, with the hope of observing subtle patterns of nonverbal marginalization in real time.

In section 2 I discuss the epistemic harms of nonverbal marginalization. Nonverbal marginalization can cause (at least) two kinds of epistemic harm: epistemic oppression and hermeneutical injustice.

Epistemic oppression involves infringement on epistemic agency, which harms the general state of social knowledge within a given epistemic community (Dotson 2012). I claim that nonverbal marginalization infringes on the epistemic agency of historically oppressed people in fundamental respects. In particular, pervasive patterns of nonverbal marginalization impair peoples' capacities to acquire and produce knowledge within their epistemic communities. For example, receiving fewer positive and affirming nonverbal cues in virtue of one's comparatively low power status infringes on one's ability to participate in knowledge production because nonverbally marginalized individuals are too intimidated to speak, ask questions, etc.

Nonverbal marginalization can also involve hermeneutical injustice (i.e., lacking the relevant concept to make sense of your experience of harm; see Fricker 2007). I argue that nonverbally marginalizing experiences are often compounded by hermeneutical injustice because victims of nonverbal marginalization lack the hermeneutical resource necessary to make sense of these experiences. I propose that the missing hermeneutical resource is the 'nonverbal marginalization' concept. Thus, there's a sense in which this project attempts to address the hermeneutical injustice by proposing a new hermeneutical resource: nonverbal marginalization.

In section three, I argue that nonverbal marginalization sheds novel light on two significant bodies of empirical literature from social psychology: imposter syndrome and performance gaps between high and low power social groups.

Regarding imposter syndrome (c.f. Clance & Imes 1978 and Bravata et al. 2020), I claim that nonverbal marginalization compounds marginalized peoples' experiences of imposter syndrome by subtly giving them evidence for their imposter attitudes. For example, people not looking at women in a business meeting subtly communicates to them that they are not valuable or central to the intellectual community of the office—in short that they are imposters. This challenges the received view of imposter syndrome in psychiatry, which assumes that imposter attitudes are triggered by agents' knowledge that identity prejudices exist in the world but are not necessarily directly caused by prejudice in their immediate environments.

Finally, nonverbal marginalization can help explain the performance gaps between high and low power social groups (e.g., men and white people outperforming women and people of color on professional and academic assessments—c.f., Mendoza-Denton 2014; Salehi et al. 2019; Shockley 2021). These results have often been explained by appeal to innate ability differences between groups (e.g., Jessim, Crawford, & Rubinstein 2015). However, I demonstrate how nonverbal marginalization can shed novel light on certain types of performance gaps, without problematic appeal to innate ability differences. Specifically, I claim that nonverbal behavior (either marginalizing or validating) affects performance in many domains. To illustrate, think about the experience delivering the same talk to a nonverbally engaged audience (exhibiting positive nonverbal behaviors like nodding and smiling) vs a nonverbally disengaged audience (exhibiting negative nonverbal behaviors like frowning, looking at their phones, and staring into space). The positive nonverbal feedback from the first audience will almost certainly translate into a better talk performance. However, receiving positive or negative nonverbal cues obviously doesn't alter your underlying ability. It's just that receiving positive nonverbal cues makes you feel more confident, so you end up giving a better talk. Thus, understanding the relationship between nonverbal marginalization and task performance can help us explain certain performance gaps between high and low power social groups while resisting empirically and socially questionable innate ability explanations.

I conclude with a discussion of how individuals and institutions can go about mitigating the effects of nonverbal marginalization, making the empirical case that we can over time habituate more equitable nonverbal behaviors. I cash this out by engaging with empirical literature on habituation, which suggests that online conscious monitoring can affect gradual changes in patterns of spontaneous nonverbal behaviors (Masala & Weber 2016; Devine et al. 2012). In doing so I demonstrate how the development of new hermeneutical resources can facilitate empirically-grounded bias intervention strategies within individual and institutional contexts.

Ballarini, Cristina (New York University) and Eric Mandelbaum (CUNY Graduate Center)

Doxastic Cartography

1. Introduction

Belief is a central concept in philosophical theorizing. Yet it's unclear to what extent theories of belief are even in conflict with one another. Dispositionalism and representational realism are in some conflict [1], as are analytical- and psycho-functionalism [2]. But is there any natural pairing between these sets? What do any of these theories have to say about the idea that beliefs "aim at truth" [3], the distinction between full belief and credence [4], the Principle of Charity [5], or the slogan that "belief is weak" [6]?

And are philosophical models of belief compatible with Heiderian theories of attitudes [7], the Elaboration Likelihood Model of persuasion [8], or the Spinozan theory of belief acquisition [9]?

Our goal, in this talk, is to map out the landscape for theories of belief. The first part of the talk groups theories of belief into four main clusters, distinguished by the different concerns around which they were developed. We outline what we take to be the main within-cluster disagreements, and suggest that there's surprisingly little disagreement between clusters. The second part of the talk then turns to some points of agreement. We argue that, despite their very different starting points and concerns, there's nonetheless a surprising amount of convergence among theories of belief.

2. Sorting Theories Into Clusters

Theories of belief have been developed in response to a number of different concerns. We group these concerns into four major clusters. Along the way, we highlight a few within-cluster disagreements, and argue for the compatibility of various theories between-clusters.

Cluster A: Explaining Behavior

Basically everyone agrees that beliefs play some role in explaining behavior. The first cluster of theories are concerned with understanding what beliefs must be like if they are to play this role. Theories diverge depending on whether they're "realists" (e.g., Fodor) or "superficialists" (e.g., Schwitzgebel) about explanations of behavior that appeal to belief, and on whether they're primarily interested in folk psychological or scientific explanations. Theorists in Cluster A include eliminativists, dispositionalists, analytical-and psycho-functionalists, interpretivists, and cognitive scientists studying topics such as attitudes, placebo effects and dissonance [10]. Here, theories of the metaphysics of belief are molded around the idea that beliefs are those mental states that interact with motivational states to result in action.

Cluster B: Content

A different source of interest in belief stems from the familiar idea that belief is a propositional attitude. The second cluster of theories have been developed alongside various theories of propositions; divergence hinges on whether theorists are committed to beliefs having relatively coarse- or fine-grained contents. Here, metaphysical and descriptive pictures of belief are molded to fit one's favored theory of propositional content. Theorists in Cluster B include Stalnakerians, content-skeptical eliminativists, as well as a number of formal epistemologists and philosophers of language [11].

Cluster C: Belief Ascriptions

Linguists and philosophers of language have long been interested in the semantics of belief ascriptions. A third cluster of theories we consider make claims about what beliefs are on the basis of how we talk about them. Issues that arise touch on topics outside of semantics, including Moore's paradox and the claim that "belief is weak". Theorists in cluster C include cognitive scientists focused on linguistics (e.g. Ray Jackendoff) formal semanticists (Kai Von Fintel), and formally oriented philosophers of language and logic (Anthony Gillies, Daniel Rothschild, Matt Mandelkern, Jeremy Goodman, and Ben Holguin) [12].

Cluster D: Epistemology

The final cluster of theories center on the idea that beliefs are the primary targets of epistemic appraisal. Theories, here, fall into one of a number of sub-clusters, depending on the aspects of epistemological practice with which they are concerned. Some theorists are interested in the assumption that we have some kind of control over what we believe [13], or the assumption that we have a special kind of access to what we believe [14]. Others are interested in the idea that beliefs "aim" at truth [15] or that they play a particular role in reasoning and inquiry [16]. And yet others are interested in making sense of the idea that beliefs come in degrees [17].

3. Common ground: Current State of Agreement for the Theories of Belief

In spite of their unaligned concerns, there's a surprising amount of agreement between extant theories of belief. Points of agreement we focus on include: The claim that beliefs play some role in explaining behavior, that belief attribution is subject to a principle of charity, that the state of believing is not luminous, and that beliefs are inferentially promiscuous. Of course, there are theorists who reject each of these claims. Still, convergence toward them–especially from such disparate starting points– provides sturdy evidence for their truth.

References

[1] Quilty-Dunn, J. & Mandelbaum, E. (2018). "Against dispositionalism: belief in cognitive science." *Philosophical Studies* 175(9), 2353-2372.

[2] Braddon-Mitchell, David, and Frank Jackson. "The Divide and Conquer Path to Analytical Functionalism." *Philosophical*

Topics 26, no. 1/2 (1999): 71–88.; Porot, N., & Mandelbaum, E. (2021). "The science of belief: A progress report." *Wiley Interdisciplinary Reviews: Cognitive Science*, 12(2), e1539.

[3] Shah, N., & Velleman, J. D. (2005). "Doxastic deliberation." *The Philosophical Review*, 114(4), 497-534.

[4] Jackson, E. G. (2020). "The Relationship Between Belief and Credence." *Philosophy Compass* 15 (6):1–13.

[5] Lewis, D. (1974). "Radical interpretation." *Synthese* 27 (July-August), 331-344.

[6] Hawthorne, J.; Rothschild, D. & Spectre, L. (2016). "Belief is weak." *Philosophical Studies* 173 (5), 1393-1404.

[7] Heider, F. (1946). "Attitudes and cognitive organization." *The Journal of psychology*, 21(1), 107-112.

[8] Cacioppo, John T., and Richard E. Petty. "The elaboration likelihood model of persuasion." *ACR North American Advances* (1984).

[9] Mandelbaum, E. (2014). Thinking is believing. Inquiry, 57(1), 55-96.

[10] Some sample theorists: Eliminativism: Churchland, P. M. (1981). "Eliminative materialism and propositional attitudes." *The Journal of Philosophy*, 78(2), 67-90.
Dispositionalism: Schwitzgebel, E. (2013). "A dispositional approach to attitudes: Thinking outside of the belief box." In *New Essays on Belief* (pp. 75-99). Palgrave Macmillan, London.
Instrumentalism: Dennett, D. C. (1991). "Real patterns." *The Journal of Philosophy*, 88(1), 27-51.
Attitude theorists: Gawronski, B., & Bodenhausen, G. V. (2006). "Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change." *Psychological Bulletin*, 132(5), 692.
Persuasion theorists: Petty, R. E., Wegener, D. T., Fabrigar, L. R., Priester, J. R., & Cacioppo, J. T. (1993). "Conceptual and methodological issues in the elaboration likelihood model of persuasion: A reply to the Michigan State critics: Specifying the ELM." *Communication Theory*, 3(4), 336-342.

[11] See, e.g., Stalnaker, R. (1984). *Inquiry*. Cambridge: MIT Press; Dretske, F. (1991). *Explaining Behavior: Reasons in a World of Causes*. Cambridge: MIT press.; Moss, S. (2015). "On the semantics and pragmatics of epistemic vocabulary." *Semantics and Pragmatics*, 8, 5-1; Moss, S. (2018). *Probabilistic Knowledge*. Oxford University Press; Yalcin, S. (2018). "Belief as question-sensitive." *Philosophy and Phenomenological Research*, 97(1), 23-47; Ramsey, W., Stich, S., & Garan, J. (1990). "Connectionism, eliminativism, and the future of folk psychology." In *Philosophy, Mind, and Cognitive Inquiry* (pp. 117-144). Springer, Dordrecht; Schwitzgebel, E. (2002). "A phenomenal, dispositional account of belief." *Noûs*, 36(2), 249-275

[12] See e.g., Hawthorne, J., Rothschild, D., & Spectre, L. (2016). "Belief is weak." *Philosophical Studies,* 173(5), 1393-1404; Goodman, J., & Holguín, B. (2022). "Thinking and being sure." *Philosophy and Phenomenological Research*; Holguín, B. (2022) "Thinking, Believing, Guessing." *Philosophers Imprint* An Von Fintel, K., & Gillies, A. (2007). "An opinionated guide to epistemic modality." *Oxford studies in epistemo*logy, 2, 32-62; Jackendoff, Ray. "The Cognitive Science of 'Belief '(The Word)." *The Cognitive Science of Belief: A Multidisciplinary Approach*(2022): 33-54.

[13] Hieronymi, P. (2008). "Responsibility for believing." *Synthese*, 161(3), 357-373; Hunter, D. (2022). *On Believing: being right in a world of possibilities*. Oxford University Press; Boyle, M. (2009). "Active belief." *Canadian Journal of Philosophy*, 39(sup1), 119-147

[14] Moran, R. (2012). *Authority and estrangement: An essay on self-knowledge*. Princeton University Press; Keeling, S. (2021). "Knowing our Reasons: Distinctive Self-Knowledge of Why We Hold Our Attitudes and Perform Actions." *Philosophy and Phenomenological Research*, 102(2), 318-341; Peacocke, C. (1998). "Conscious attitudes, attention, and self-knowledge. In *Knowing our own minds*, 63-98.

[15] Shah, N., & Velleman, J. D. (2005). Doxastic deliberation. The Philosophical Review, 114(4), 497-534; Helton, G. (2020). "If you can't change what you believe, you don't believe it." *Noûs*, 54(3), 501-526.; Flores Henrique, C. (2022) "Bad believers: evidence-resistance, rational persuasion, and social change (Doctoral dissertation, Rutgers University-School of Graduate Studies); Glüer, K., & Wikforss, Å. (2013). "Against belief normativity." *The aim of belief*, 80-99.

[16] Boghossian, P. (2014). "What is inference?" *Philosophical studies*, 169(1), 1-18. Friedman, J. (2019). "Inquiry and belief." *Noûs*, 53(2), 296-315.

[17] F.P. Ramsey (1926) "Truth and Probability", in Ramsey, 1931, *The Foundations of Mathematics and other Logical Essays*, Ch. VII, p.156-198, edited by R.B. Braithwaite, London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company. 1999 electronic edition.; Jeffrey, R. (1970). "Dracula Meets Wolfman: Acceptance vs. Partial Belief." In Marshall Swain (ed.), *Induction, Acceptance, and Rational Belief*. D. Reidel. pp. 157-185; Eriksson, L., & Hájek, A. (2007). "What are degrees of belief?" *Studia Logica*, 86(2), 183-213.; Icard, T. (2016). "Subjective probability as sampling propensity." *Review of Philosophy and Psychology*, 7(4), 863-903.

Barner, Alma (University of Salzburg)

Reasoning with Synaesthesia

In this paper I aim to present a realistic account of some of the epistemic roles synaesthesia plays in ordinary reasoning. I argue in particular for two claims: i) synaesthetic experiences can justify beliefs about the world and ii) synaesthetic experiences play an epistemic role in justifying beliefs about the past.

Seeing the letter 'A' triggers a redness sensation and smelling lavender triggers an experience of soft, three-dimensional tubes. These are examples of synaesthetic experiences. In the last decades, we have witnessed a steady increase in our understanding of the nature of this phenomenon, yet so far there has been very little work on carving out possible cognitive or, more specifically, epistemic roles synaesthesia plays. This is somewhat surprising, as among synaesthetes themselves, synaesthesia is generally perceived to be a positive feature of their lives. Moreover, subjective reports highlight various uses that synaesthetic experiences are commonly put to in ordinary reasoning. In particular, it is well-documented that synaesthetes perform well above average on a range of memory tasks (see in particular Rothen et al. (2012)). One possible explanation for this is that synaesthetic experiences play an epistemic role in justifying beliefs about the past.

From a philosophical perspective, synaesthetic experiences present us with what seems to be an intractable epistemic puzzle: They are highly individual, seemingly arbitrary associations that are not in the business of informing us about the world. Consider grapheme-color synaesthesia as a classic example: Subjects robustly and involuntarily associate color imagery with graphemes. Upon seeing the letter 'P' a subject cannot help to conjure up a sensation of the color purple. How could such a sensation play any epistemic role in reasoning? What types of beliefs would it justify? How could it justify beliefs about the external world, for example?

In the first, introductory part of the paper, I briefly outline my understanding of the nature of synaesthetic experiences. When synaesthesia first captured the interest of researchers, it was defined as 'the stimulation of one sensory modality reliably caus[ing] a perception in one or more different senses' (Cytowic, 1997). Nowadays this definition is considered imprecise at best. First, as synaesthesia is a highly heterogeneous condition, it evades short definitions like the one given. Standard definitions typically do not capture all types of synaesthetic experiences. Second, it is controversial whether synaesthetic experiences are best classified as perceptual or rather imagistic instead. In fact, I endorse the following view, which is most explicitly advanced by Nanay (2021): synaesthetic experiences are typically constituted by multimodal imagery. As an example, in my view, grapheme-color synaesthesia is constituted by imagery. Seeing the letter 'A' triggers a mental image of redness. In this first part of the paper I give reasons in favor of this view.

In the second part of the paper I briefly discuss and reject philosophical accounts of the epistemic value of synaesthesia. Existing proposals (cf. Matthen (2017), Sollberger (2013), (2017)) center around the question whether synaesthetic experiences can be veridical. Matthen (2017) argues that, like ordinary perceptual experiences, synaesthetic experiences can provide immediate, non-inferential justification for beliefs about the presence of their inducers. Sollberger (2017) claims that they can merely provide such justification on the basis of an inference. I argue that both accounts fail to do justice to a range of empirical data on constitutive properties of synaesthesia, such as, among others, the nature of the synaesthetic experience and its inducer. I also show that their views on the epistemic value of synaesthesia are mistaken in part due to the fact that they see synaesthesia as a purely perceptual phenomenon without doing justice to the role that mental imagery plays.

In the third part of the paper, I present a more realistic account of different epistemic roles synaesthesia might play. As a preliminary, I offer some empirical evidence for such roles in the form of self-reports. These highlight a broad range of different usages of synaesthetic experiences in ordinary reasoning, depending on the type of synaesthesia. This suggests that a part of the solution to the epistemic puzzle is to take at face value the heterogeneous nature of synaesthetic experiences. I demonstrate this diversity by providing an in-depth analysis of three types of synaesthesia:

First, I argue that grapheme-color synaesthesia can play an inferential role in forming beliefs about the past. Second, I argue that time-space synaesthesia can justify beliefs about time. As I explain, this type of synaesthesia is – in typical instances – constrained imagination in the sense of Kind (2016), having epistemic status in the

sense of Myers (2020)). More specifically, what is represented in an experience of time-space synaesthesia is constrained by the subjects' dispositional beliefs about time. Third, I argue that there is currently not sufficient evidence to warrant the claim that mirror-touch synaesthesia enables subjects introspective access to what it is like for someone else to be touched. More generally, I give reasons to question the intuitively plausible thought that there is an epistemic link between this type of synaesthesia and empathetic beliefs.

References

Cytowic, R. E. (1997). Synaesthesia: Phenomenology and neuropsychology—A review of current knowledge. In S. Baron-Cohen & J. E. Harrison (Eds.), *Synaesthesia: Classic and contemporary readings*. Blackwell Publishing. 17–39.

Kind, A. 2016. Imagining Under Constraints. In: Amy Kind and Peter Kung (eds.): *Knowledge Through Imagination*. Oxford University Press. 145–159.

Myers, J. 2021. The Epistemic Status of the Imagination. *Philosophical Studies* 178 (10): 3251–3270.

Matthen, M. 2017. When is synaesthesia perception? In: Ophelia Deroy (ed.): *Sensory Blending: On Synaesthesia and Related Phenomena*. Oxford, GB. Oxford University Press. 166–178.

Nanay, B. 2021. Synaesthesia as (multimodal) mental imagery. *Multisensory Research* 34 281–296.

Rothen, N., Maier, B. & Ward, J. 2012. Enhanced memory ability: Insights from synaesthesia. *Neuroscience & Biobehavioral Reviews*, 36: 8. 1952-1963.

Sollberger, M. Rethinking synaesthesia. *Philosophical Psychology* 26 (2).

Sollberger, M. 2017. Can Synaesthesia Present the World as it Really is? In: Ophelia Deroy (ed.): *Sensory Blending: On Synaesthesia and Related Phenomena*. Oxford, GB. Oxford University Press. 179–188.

Battich, Lucas (École Normale Supérieure, Paris Sciences et Lettres University)

Shared perception and the sense of reality

When we perceive the world, we experience a host of perceptual contents. We see shapes, colours, and movement, experience textures through touch, and hear the timbre and frequency of sounds. But there is more to perception than sensory contents: we also have the feeling that the things we perceive are real and not, for example, merely imagined. We can differentiate veridical perceptual experiences from imagination, because veridical perception comes together with a feeling that we are perceiving the real world. We have a sense of reality. This sense is often taken for granted, but what makes an experience feel real? What gives a sense of reality to our perceptual experiences?

Most scientific and philosophical treatments of the topic take an individualistic approach. Many perceptual experiences, however, are not merely individual, but socially shared: infants learn to look where others are looking, doctors jointly assess radiographs to detect tumours, hunters track prey together, and musicians will jointly experience the music they play. In this paper, I examine whether shared perception is a normative precondition to the possibility of experiencing the world as real. Can we have a sense of reality without the ability to coordinate and share our perceptions with others? In other words, can we have a sense of reality if we have never experienced sharing percepts with others?

Several factors have been proposed as causal determinants of the SoR in perception, including the vividness of the perceptual experience, the lack of voluntary control in genuine perception compared to typical cases of mental imagery, and the combination of sensory information from different modalities (Deroy & Rappe, 2022; Dijkstra et al., 2021; Farkas, 2014). But neither of these features is necessary: we often have the sense that what we perceive is real, even when it may be fuzzy, unclear, and perceived only from one modality; and the feeling of voluntary control can equally be lacking in imagery that also lacks a sense of reality. The evidential support provided by shared perception towards the same object is obviously not necessary, as we perceive the world as real even when there is no one else nearby. But there may be normative preconditions for experiencing the sense of reality in perception. Is shared perception a normative precondition to the possibility of experiencing the world as real?

Several analytic philosophers argue that the ability to share our perceptual experiences with others is, in fact, essential for the concept of a shared objective world, where mind-independent objects are attended in common (Davidson, 1999; Seemann, 2019). Similarly, in the phenomenological tradition,

Husserl argued that intersubjective experience is part of what constitutes the very idea of an objective world: it is a condition of possibility for any knowledge of external objects (Husserl, 1977). Without social engagement, there is no concept of objectivity.

These views are concerned with the normative preconditions for the concept of an external objective reality. The sense of reality, however, is a basic feature of our perceptual experience, which may be shared by pre-linguistic and non-conceptual populations, including infants and non-human animals. We need a reevaluation of this tradition to address the sense of reality as a non-conceptual feature of perceptual experience. Here, I will analyse the stronger position that sharing percepts with others is a normative precondition to the sense of reality. This is the "intersubjective sense of reality" claim: perceptual reality is socially mediated at the basic level of phenomenal experience.

First, I will go consider which candidate factors may be normatively necessary for the possibility of distinguishing self-generated versus externally triggered activity in perception. From an individualist approach, the idea of a minimal sense of self is one of the strongest candidate factors as a precondition for the possibility of distinguishing self-generated versus externally triggered activity in perception. The most elaborate analysis of a minimal sense of self in action and perception draws from the mechanisms necessary for guiding adaptive behaviour (Forch & Hamker, 2021; Gallagher, 2000; Pacherie, 2011). All but the most primitive animal species regulate their stream of sensory information together with their own goal-directed movements (Fuster, 2004). At the very least, this regulation requires the ability to separate two different causes for changes in sensory signals: those caused by the external environment, and those caused by the agent itself as it moves and acts through its environment. This basic ability already provides the organism with a rudimentary sense of self and of agency that can in principle be sufficient to allow the development of a perceptual sense of reality.

This approach has its limitations, however. Next, based on current work in theoretical psychology, I categorise different levels of sharing and coordinating percepts. From this, I argue that the sense of reality in perception is not an all-or-nothing component of experience but a phenomenon with different levels of complexity, where the ability to coordinate my perceptual attention to an object together with another individual in a flexible manner is a crucial component. On this proposal, the sense of reality is more sophisticated (or more efficient, better at discriminating between imagery and perception) in an agent that has the ability to be sensitive to others' perceptual perspectives.

Finally, I close by considering two important empirical corollaries of this proposal in psychiatry and our understanding of imagination and hallucination in non-human animals and artificial agents. The social approach to the sense of reality will give conceptual clarity to research on hallucinations in schizophrenia. This approach also has consequences for the understanding of imaginative abilities in artificial intelligent agents and non-human animals: creatures that lack the ability to coordinate perception with others, may also lack the ability to separate imaginings and hallucination from objective reality.

Beckman, Elizabeth (University of Michigan)

Moral responsibility for empathic failures: a response to David Shoemaker

Can we be morally responsible for failing to empathize with one another? If so, how do we characterize those failures and set conditions for responsibility? This paper serves as both a response to David Shoemaker's account of "empathic control," and a general exploration of the former two questions. I argue that Shoemaker's account, while a helpful start, is ultimately unsuccessful because he 1) relies on an overaccommodation of folk attitudes, 2) a tenuous account of "empathic control" and corresponding capacities 3) a problematic omission of the epistemic condition of moral responsibility, and 4) he neglects the role of agent-relevant information. In the end, I hope to draw attention to difficulties in sketching any such account of moral responsibility for empathic failures.

The first section of the paper is dedicated to a preliminary exploration of empathy as a concept. Some psychologists have suggested that empathizing might involve the suppression of an automatically

generated self-perspective (that is presumably formed via simulation or theorizing) (Decety, 2004). On this picture, many empathic failures, so to speak, arise from a failure to deploy inhibitory control. Now what kinds of empathic interactions or failures are the target of analysis here?

In order to frame our discussion, I use the following example from Shoemaker:

"Ayisha comes home from a work meeting visibly upset. Her spouse Rahul asks what is wrong, and she tells him: Her boss didn't consult the opinion of any women at the meeting, asking only the men what they thought. Rahul calmly and efficiently starts counseling Ayisha on what she should do in response: She needs to send her boss an angry email, and then contact his superior, and then rally the other women to send him a list of demands, and so forth. Ayisha stops Rahul in his tracks, saying, "I just want you to understand and appreciate what it was like for me, sitting in that meeting, not being acknowledged. I don't want you to 'solve the problem' or tell me what to do; I know what to do." So Rahul imagines himself at her meeting, and feels himself getting all riled up with righteous anger at the injustice of being ignored, and still only sees reasons to right the wrong, to resolve the problem. Thus, in response to Ayisha's continuing demand that he simply understand and appreciate what it was like for her, Rahul responds: "But how will that solve anything?"

(Shoemaker, 2022, p. 101-102)

Shoemaker assumes that our intuitions side with finding Rahul responsible for such an empathic failure and argues that, in response, we should find ways of accommodating these intuitions. I dispute this move on both accounts: one, I think our intuitions are responding to Rahul's refusal to acquiesce to Ayisha's request (which remains independent from details on empathic capacities) and two, conditions for moral responsibility should not simply be mirrors for current blaming practices. If this is true, where would we start in determining what features of this case license us to find the agent, Rahul, responsible?

In order to be morally responsible for our actions, many philosophers stipulate that we should 1) be aware of what we are doing, i.e. satisfy an epistemic condition and 2) have control over those actions, i.e. satisfy a control condition. One of my interventions here is to suggest that many empathic failures (regarding inhibitory control) concern the satisfaction of the former epistemic condition. If I fail to empathize, or grasp the experience of the other, I may have failed to suppress my self-perspective. Suppression, as a regulation strategy, is not presented to me as 'the thing to do' – I may not know to deploy inhibitory control, for instance (and then of course then I fail to know the other in some respects).

Shoemaker solely focuses on the control condition of moral responsibility. Situations in which we "fail to register [a] fact" (p. 97), like Rahul does with Ayisha, constitute cases of "empathic disregard." He believes that we can be criticizable certain kinds of perceptual failures, like only seeing the duck side of the famous duck-rabbit image. He argues that, in both the duck-rabbit case and Ayisha/Rahul case, the agent is criticizable on account of having control over the way a situation appears to them. I criticize this move for two reasons: one, we can't simply posit without justification that Rahul, or any agent, has a functioning mechanism of sorts and just does not use it and two, the epistemic condition seems to be a more natural place to localize this discussion. Failing to register a fact seems like a clear failure of the epistemic condition.

I argue that an account of moral responsibility for empathic failures demands a bit more mechanistic clarity. What exactly are we holding the agent responsible for? I argue that we need to be careful about when we posit functioning versus broken mechanisms. We can't simply assign them where our intuitions lead us – Shoemaker supposes that we find Rahul responsible because he fails to use a functioning mechanism (whereas he suggests we might excuse an unempathetic nurse on account of having a broken or worn-down mechanism). What sort of mechanism exists within in an agent remains an independent psychological fact that should be established separately.

Overall, I caution any approach in which we build an account of responsibility in response to folk attitudes. This motivation alone likely accounts for the missteps in Shoemaker's account. In this paper, I dispute the existence of "empathic control" in the way Shoemaker has described it and I highlight the relevance of the epistemic condition of moral responsibility. In the end I do not suggest we simply excuse agents like Rahul, but that we exercise caution and modesty in these types of judgments. It may turn out that information we need to hold an agent responsible (what goals they have, what the status is of their "empathic mechanisms") render us unqualified to issue judgments of moral responsibility.

Berger, Larissa (University of Siegen)

Moral Perception, Recalcitrance, and the Influence of Moral Beliefs

Suppose you discover a group of kids burning a cat. Most of us will have an immediate awareness that what the kids are doing is wrong. Some philosophers suggest that this awareness is perceptual, that is, that we can have perceptual experiences of moral properties, some of which are veridical. I will hypothetically assume that this thesis of moral perception is true. My question will be: how does moral perception relate to (moral) concepts and beliefs?

My starting point will be the phenomenon of recalcitrant perception. Consider the Müller-Lyer illusion. The two lines visually appear to be of different lengths, even though you know that they are in fact of the same length. And this conflict persists. I suggest the following definition of 'recalcitrant perception' (RP):

A perceptual experience is recalcitrant iff

- it comprises an illusionary part,
- the illusionary part conflicts with a true belief held by the perceiver,
- and the illusionary part constantly persists (i.e., cannot be adjusted to the true belief).

RP has been taken to support the theses that (a) some perceptual content is non-conceptual and (b) that some perceptual content is cognitively impenetrable. Crane (1992) argues that some perceptual content is non-conceptual (a) where a state with non-conceptual content does not necessarily involve concepts. For Crane, concepts are "the inferentially relevant constituents of intentional states." (Crane 1992, 147) He conceives of three kinds of inferential relations of states: evidential, deductive, and semantic relations. RP shows that (some) perceptions do not stand in evidential relations. Crane takes this to support his thesis that perceptual content is non-conceptual. (Since perceptual content could still figure in evidential or semantic relations, RP does not provide conclusive evidence for this thesis.) Pylyshyn takes RP to support the thesis that some perceptual content, viz., that associated with early vision, is cognitively impenetrable (b). If a system is cognitively penetrable, then "it can be altered in a way that bears some logical relation to what the person knows," beliefs, desires, or expects (Pylyshyn 1999, 343). Since in the case of RP the perception cannot be altered, RP is supposed to support the thesis of cognitive impenetrability.

In the literature on moral perception (MP), it is sometimes assumed that MP can also be recalcitrant (see, for example, Cowan 2015, 174 f.). A first candidate for recalcitrant MP is to be found in fictional cases (e.g., movies). Here, MP would be "persisting in the face of knowledge that there are no ethical properties instantiated." (Cowan 2015, 174 f.) Three reasons speak against qualifying such fictional cases as recalcitrant MP: First, the conflict is confused, since the perceiver believes and sees that, for example, the action in the movie is wrong. Second, illusions are to be detached from fiction. In fictional cases, people usually don't have the illusion of the scene being real. Third, given a representational theory of perception, illusions concern perceptual content, but it is highly disputed whether what is called the 'sense of reality' concerns perceptual content. More promising candidates for recalcitrant MP are cases of change in moral belief. Imagine that Mary was brought up being told that homosexuality is morally wrong. Later she comes to believe that homosexuality is morally permissible. Nevertheless, whenever Mary encounters a homosexual couple in the street, she perceives it as morally wrong. Here, the perceptual experience comprises an illusionary part which conflicts with a true (moral) belief. However, the illusionary part does not (always) persist constantly. Although it may take some time, Mary can come to unsee the homosexual couple as morally wrong (see similarly Väyrynen 2018, 125). The fact that the adaption of the moral perceptual content to the moral belief takes some time can be explained by the lack of objective measurement, by the fact that the acquisition of expert seeing usually takes time, and by the fact that there may be other mental states penetrating the perceptual content (e.g., emotions). In a third case of supposed recalcitrant MP there is no change in moral belief, but in qualifying an action as falling under a moral belief. Suppose Max watches Anna stabbing a knife into somebody's heart. Max holds the belief that murder is morally wrong and he classifies Anna's action as murder. He learns that Anna is a heart surgeon and, indeed, performs a surgery to save a patient's life. Still, Max cannot help but see Anna's action as morally wrong. This is also not an example of recalcitrant MP. Again, after some time Max can come to unsee Anna's action as morally wrong.

My diagnosis is that MP cannot be recalcitrant. What does this imply for a theory of MP? Drawing on Crane's theoretical framework, we can say that if a kind of perception can be revised in the light of

counterevidence, it stands in evidential relations. And, if a mental state stands in evidential relations, it will count as conceptual. So we can conclude that MP has conceptual content. Similarly, we can conclude that MP is prone to cognitive penetration understood as a non-trivial influence from cognitive states such as beliefs. According to Macpherson, however, there are two kinds of cognitive penetration: for classic cognitive penetration, it is necessary that the perceptual experience is causally affected by a certain propositional attitude in order to yield the particular perceptual content; for cognitive penetration lite, such a causal influence can and does occur but is not necessary to yield the particular perceptual content (cf. Macpherson 2015). From the fact that there is no recalcitrant MP we can merely conclude that MP is prone to cognitive penetration lite.

Understanding MP on a cognitive penetration model comes with crucial consequences: First, it opens the way for epistemic downgrade where too much weight is given to the penetrating states, or the latter are simply unjustified (cf. Siegel 2017). But second, it also allows for perceptual improvement by concept-dependent expertise which leads to subjects perceiving more accurately within the moral domain or being more sensitive to moral properties (cf. Stokes 2021).

Bibliography

Cowan, Robert (2015a), "Perceptual Intuitionism", in: *Philosophy and Phenomenological Research*, 90 (1), 164–193.

Crane, Tim (1992), "The Nonconceptual Content of Experience", in: T. Crane (Hrsg.), *The Contents of Experience: Essays on Perception*, Cambridge: CUP, 136–157.

Macpherson, Fiona (2015), "Cognitive Penetration and Nonconceptual Content", in: J. Zeimbekis & A. Raftopoulos (eds.), *The Cognitive Penetrability of Perception. New Philosophical Perspectives*, Oxford: OUP, 331–358.

Pylyshyn, Zenon (1999), "Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception", in: *Behavioral and Bran Sciences*, 22 (3), 341–423.

Siegel, Susanna (2017), *The Rationality of Perception*, Oxford: OUP.

Stokes, Dustin 2021, *Thinking and Perceiving*, London & New York: Routledge.

Väyrynen, Pekka (2018), "Doubts about Moral Perception", in: A. Bergqvist & R. Cowan (eds.), *Evaluative Perception*, OUP, 109–28.

Blakey, Kirsten (University of Stirling) and Giacomo Melis (University of Stirling)

Reimagining metacognition: The transition from unreflective to reflective epistemic thinking

The distinction between object level cognition and metacognition has been the subject of fierce debate among developmental and comparative psychologists over the past few decades. Some would describe metacognition as "thoughts about other thoughts" or as metarepresentational (only occurring when a mental representation represents another mental representation; Carruthers, 2018; Perner, 2012). However, this metarepresentational account hinders the study of metacognition in non-linguistic populations such as infants and non-human animals. At least in part driven by the ambition to account for the possibility of metacognition in non-verbal subjects, others defend a proceduralist view according to which mental processes and states involved in controlling and monitoring one's cognitive activities are metacognitive regardless of whether they involve metarepresentation (Proust, 2019).

The two approaches outlined have focused primarily on the nature of metacognitive states, though they have said little about the transition from the object level to the meta-level. Moreover, they sometimes talk past each other by focusing differently on conscious (metarepresentationalist) and non-conscious states (proceduralists). We outline an alternative approach which promises to explain the transition from object level to the meta-level, while vindicating some of the main insights of the two rival views. Like metarepresentationalists, we focus on conscious thought but, in line with proceduralists, we allow for non-metarepresentational forms of metacognition.

We take an interdisciplinary approach engaging developmental and comparative psychology, epistemology and the theory of reasons, which suggest that the cognition/metacognition opposition may be characterised—at least when the focus is on conscious thought—as the one between unreflective and

reflective responsiveness to evidence or reasons. The framework we outline is compatible with view that non-linguistic creatures may be reflective and sets out a series of levels which describe both unreflective and reflective thinking.

At the unreflective level we describe first-order epistemic thinking, which includes the formation of beliefs based on (positive) evidence, and their revision based on counterevidence (overriding defeaters, e.g., belief <P> is replaced with its negation <not P>). This type of belief revision remains unreflective, as the subject has a mental representation of the content of their belief or its negation, but they do not reflect. That is, they do not (or cannot) individuate the evidence for the belief, nor do they assess it as such.

The transition to the next level of higher-order epistemic thinking, or basic reflective thinking, emerges from first-order epistemic thinking through generalisation. Experience of several instances of overriding defeaters coming from the same source may lead an agent to process an undermining defeater such as <the source of evidence is unreliable> and respond by suspending judgement. While the overriding defeaters are available at the reflective level, undermining defeaters require reflection on the reasons for the belief. However, the basic reflective thinking available at the level of higher-order epistemic thinking needn't require language, therefore a response to undermining defeaters in non-linguistic populations would be compatible with the notion that very young children and non-human animals are capable of reflective thinking.

A particular advantage of this framework is that it offers empirically testable hypotheses in non-linguistic populations. Indeed, a couple of empirical studies have recently assessed the capacity for overriding defeaters in children and great apes (O'Madagain et al., 2022), and undermining defeaters in children (Schleihauf et al., 2022). These studies offer the first suggestion that it may be possible to assess both unreflective and reflective epistemic thinking, as well as the transition from one to the other. I will discuss the implications of empirical assessment of the new framework and how it compares to the current accounts of metacognition.

Bleijlevens, Natalie (Developmental Psychology, University of Göttingen, Germany) and Tanya Behne (Developmental Psychology, University of Göttingen, Germany)

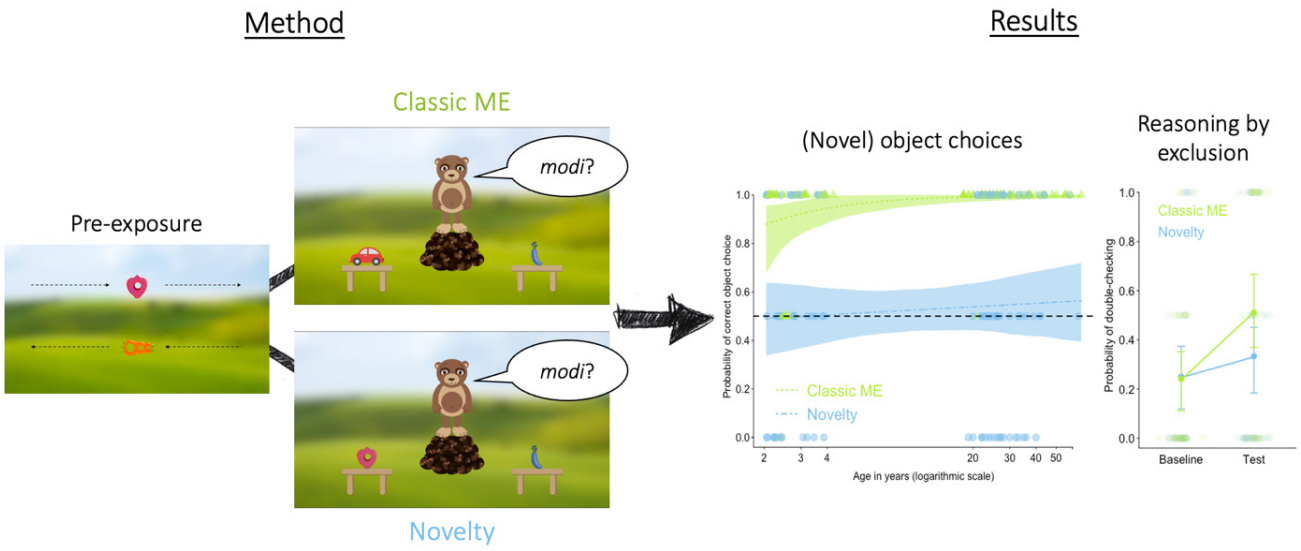
Young children and adults use reasoning by exclusion, rather than attraction to novelty, to disambiguate novel word meanings

Upon hearing a novel label, listeners tend to assume that it refers to a novel, rather than a familiar object (see Lewis et al., 2020 for a review). While this disambiguation or Mutual Exclusivity (ME) effect has been robustly shown across development, it is unclear what it involves. Do listeners use the pragmatic context (Bleijlevens et al., 2023; Bohn et al., 2022; Clark, 2015; Diesendruck & Markson, 2001) and their lexical knowledge (e.g., Lewis et al., 2020; Markman & Wachtel, 1988) to exclude the familiar object and thus select the novel one? Or is the effect, at least in early childhood, simply based on an attraction to novelty and a direct mapping of novel label to novel object (Horst et al., 2011; Mather & Plunkett, 2012)?

In a pre-registered online-study with 2- to 3-year-olds (n=75) and adults (n=112), we examined i) whether relative object novelty alone (without pragmatic or lexical information) could account for participants' disambiguation and ii) whether participants' decision processes involved reasoning-by-exclusion strategies. Participants encountered either a known and an unknown object (classic ME condition) or two unknown objects, one completely novel and one pre-exposed (novelty condition) as potential referents of a novel label. Reasoning-by-exclusion was assessed by adults' explanations and children's looking patterns ("double checks"; Halberda, 2006): Upon hearing the novel word, children were expected to switch their gaze to the distractor object before switching back to (and staying at) the target object.

Our pattern of data across measures revealed that in the classic ME condition, children and adults significantly chose the novel object and both used reasoning-by-exclusion. In contrast, in the novelty condition, children and adults chose objects randomly. Across conditions, a retention test revealed that adults remembered their prior selections, while children's performance was more fragile. These results

suggest that referent disambiguation may not be based on relative object novelty alone. Instead, to resolve referential ambiguity, both young children and adults seem to make use of pragmatic and/or lexical sources of information and to engage in reasoning-by-exclusion strategies.



References

Bleijlevens, N., Contier, F., & Behne, T. (2023). Pragmatics aid referent disambiguation and word learning in young children and adults. *Developmental Science*, e13363. https://doi.org/10.1111/desc.13363

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2022). Predicting pragmatic cue integration in adults' and children's inferences about novel word meanings. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0001216

Clark, E. V. (2015). Common Ground. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 328–353). John Wiley & Sons. https://doi.org/10.1002/9781118346136.ch15

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630–641. https://doi.org/10.1037/0012-1649.37.5.630

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*(4), 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, *118*(2), 234–244. https://doi.org/10.1016/j.cognition.2010.10.015

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, *198*, 104191. https://doi.org/10.1016/j.cognition.2020.104191

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, *36*(7), 1157–1177. https://doi.org/10.1111/j.1551-6709.2012.01239.x

Blomkvist, Andrea (London School of Economics and Political Science) and Ali Boyle (London School of Economics and Political Science)

Episodic Memory for Learning: Insights from Artificial Agents

Episodic memory is a crucial aspect of human cognition. It heavily influences our sense of identity and our decision-making, and the loss of episodic memory, as in Alzheimer's disease or episodic amnesia, has severe consequences for a person's life (Rosenbaum et al., 2005). Yet, there is little agreement in philosophy

and psychology about the function of episodic memory: what is it for? A popular view currently holds that the function of episodic memory is to enable mental time travel, that is, the simulation of past and future events (Schacter and Addis, 2007; De Brigard, 2014). In this talk, however, we will argue that episodic memory has a more fundamental function, namely, episodic memory is for learning (Gallistel & King, 2009). We provide a novel argument for this from artificial intelligence. Firstly, we give a brief overview of current AI research on episodic memory. Secondly, we argue that despite obvious implementational differences between human and AI episodic memory, they are relevantly similar on the algorithmic level, enabling us to infer the function from AI. Thirdly, we show how this methodology is fruitful as it bypasses some problems pertinent to human research, and we argue that experiments on AI support that learning is the function of episodic memory.

Firstly, we will begin with an overview of how episodic memory-like features have been added to artificial intelligence in recent years. Several recent high-profile advances in artificial intelligence have involved algorithms resembling episodic memory. For instance, the AI company DeepMind achieved a new state of the art in playing Atari using a deep reinforcement learning algorithm equipped with an 'episodic buffer', allowing it to store information about past games (Mnih et al., 2013). The popular SOAR-architecture has also added episodic memory-like features in the form of a 'snapshot' memory, a copy of working memory which is then retrievable from an episodic memory store (Nuxoll and Laird, 2012). This has enabled the AI to learn from past actions, greatly improving its performance on decision-making tasks amongst others.

Secondly, to be able to use cases of AI episodic memory in order to infer the function of human episodic memory, these systems ought to be relevantly similar. There is a glaring difference here: AI episodic memory is implemented in artificial matter, whereas human episodic memory is implemented in biological matter. However, we will argue that there are two reasons to think that this does not hinder our project. Firstly, despite being implemented in different matter, many of the AI architectures are heavily inspired by human biology and attempt to replicate human memory down to the neural level. For example, the LEABRA architecture is comprised of neural networks containing artificial neurons, and the macrostructural organisation aims to replicate the structural organisation of the human hippocampus (O'Reilly et al., 2016). Secondly, and more importantly, we will argue that even if implementation details differ, the more important point concerns how information is processed on the algorithmic level (Marr, 1982). Here, we will demonstrate distinct similarities between human and AI episodic memory. For example, the LIDA architecture includes several processes known to be present in human memory, such as encoding, storing, recalling, consolidation, and even decaying (Franklin et al., 2016). Hence, when evaluating the information processing, we show that the cognitive architectures are relevantly similar, and this justifies the functional inference we will make from AI to human episodic memory.

Thirdly, we will demonstrate how studying AI episodic memory is fruitful for our endeavour, as well as argue that doing so indicates that episodic memory is for learning. One excellent way of finding out the function of a system is by studying what difference adding or removing it makes to a system's performance on a specified set of tasks. In research on human subjects, we are limited as we can only study the ramifications of the natural loss of episodic memory or the development of episodic memory in infants and children. But it is difficult to isolate the loss of episodic memory as the sole cause for impaired performance on tasks, as other cognitive functioning is often affected too, and we cannot control for potential confounds. Further, other problems face us when studying the development of episodic memory, as it develops alongside other cognitive functions such as language, again making it difficult to pinpoint the unique contribution of episodic memory to task performance. We will argue that experiments on AI provide a way to bypass these problems, as episodic memory modules can easily be added or subtracted from AI. In fact, we already have substantial data on the effects this has on task performance, and we will argue that improved task performance in AI is best explained as a result of improved learning abilities due to the addition of episodic memory. For example, we will show that AI systems are able to learn more efficiently as episodic memory allows for offline learning via replay of experienced episodic (Nuxoll and Laird, 2012). This in turn improves a systems' performance with regards to other tasks, such as optimisation and decision-making tasks.

We will conclude that the function of episodic memory is learning, contra the current mental time travel trend. Our conclusion is supported by the results of experiments adding and subtracting episodic memory to AI systems, and we have shown that the inference from the function of AI episodic memory to the function of human episodic memory is warranted, as the systems are relevantly similar on the algorithmic level. Finally, we hope that our methodology paves a new way to learn about human episodic memory from artificial agents.

Borgoni, Cristina (University of Bayreuth)

The enjoyment of interlocution

When we are children, our interpersonal lives are massively permeated by play. When we grow up, many of our interpersonal interactions take the form of interlocution. But what is the relationship between these two robust and prevalent forms of social interactions that pervade different stages of our life as human beings? How do we pass from playful interactions to conversational interactions? In this talk, I will explore the relationship between play and communication by focusing on the empirical work on proto-conversations. While proto-conversation has been studied as a precursor of human communication, I would note it is itself one type of social play. I suggest that philosophical reflections on these early interpersonal interactions, which are fun and enjoyable in its most primitive forms, can illuminate some central aspects of mature interlocution.

The term "proto-conversation" was proposed by Bateson (1975) to identify a type of joint performance that, arguably, figures as one of the earliest social interactions between infants and adults that resemble conversation. Thus, analyzing proto-conversations might be illuminating for understanding language development (Bateson 1975: 102). According to Bateson's (1975) initial research, infants can engage in proto-conversations at as early as 2 months of age , which are typically exemplified by the scene below:

(…) we see infants and mothers gazing at each other, as each smiles and vocalizes, apparently with pleasure and a sort of delighted courtesy. As with adult conversation, there is near-constant communication in one modality (visual) and intermittent, alternating communication in another. (Bateson, 1975: 101)

The parallel between this scene and mature conversation involves several elements: in proto-conversations, mother and child are co-participants in a sustained face-to-face interaction––the typical situation of conversation––in which they respond to each other's vocalizations; the structure of mature conversation, where adults take turns, is also mirrored in the structure of proto-conversations where adult and infant generally alternate their vocalizations by taking turns (Bateson 1975: 104).

Despite the multiple studies concerning proto-conversations during the seventies (see for example the works of Stern 1977 and Trevarthen, 1979, cf. Csibra 2010), no consensus has been reached regarding the place of proto-conversation in language development as Bateson (1975) envisaged. Nevertheless, it seems to be a consensus that proto-conversations indeed share a good deal with mature conversation, and this observation is backed up by the analysis of the acoustic and temporal structure of such interactions.

Thus, for the current work, I maintain the basic intuition of Bateson (1975), according to which proto-conversation is a type of social interaction that belongs to the genus of social interactions that will later develop into linguistic conversations as one of its species. In this talk, I also propose that proto-conversation belongs to a further species within the genus of social interactions: the play species (see Stern 1974).

In the first part of the talk––after explaining the empirical work on proto-conversations––I explain some key distinctions in the philosophical and empirical literature on play and characterize proto-conversations as a type of social (in contrast with solitary) play (in contrast with games) activity (in contrast with pretending play). Also in this part, I propose a working concept of play as a type of autotelic activity that is essentially fun. This is not a general definition of play (such as the one pursued by Suits 1977) but it is one that corresponds to the characteristics of paradigmatic instances of social play in the age frame that I am analyzing, and it is also shared by many scholars of play (e.g. Suits 1977, Huizinga 1955 and Feezell 2013).

In the second part of the talk, I address the question "what exactly is preserved from playful proto-conversation through its development into interlocution?". I first construct one reply to this question based on the philosophical work of Wittgenstein, who conceives language in analogy to play and games (Egan 2013). I finally propose an alternative picture of the relationship between communication and play based on the essence of the play being fun.

In Philosophical Investigations, Wittgenstein (1953) introduces his famous notion of "language-games" to consider several aspects of language as a practice, analogous to what we do in plays and games. The analogy between communication and game concerns, among other things, the variability and plasticity of their rules (see Egan 2013). In the talk, I highlight that despite the value of such an analogy based on how regularities are constitutive of both playing and language, a further picture of their relationship can be reached by focusing on what is essential to play. When we look at infants engaging in playing episodes, part of what makes those episodes an instance of play is that they are fun. My proposal focuses on that aspect.

My proposal is thus one that pursues a continuity story between communication and social play, rather than being a mere analogy between two independent social interactions. This is a proposal that has a clear connection with the common-sense idea that interlocution can be fun. Despite playing and interlocution standing as two independent forms of social interaction, I propose that the fun character of interlocution, when it features it, is inherited from its playful ancestor found in our early infancy.

Brozzo, Chiara (Universitat de Barcelona) and Joshua Shepherd (Carleton University)

The principle of least cognitive effort: an underexplored driver of action

It seems obvious that, in a mundane sense, agents are generally in control of their behavior. This is because much of an agent's behavior is in accordance with the agent's desires, goals, and intentions. Exceptions come in for analysis, of course – things like action slips, deviant causation, and weakness of will. But the fact that these stand out as exceptions in the philosophy of action suggests the pervasiveness of agentive control.

In this paper we point out that a powerful explanatory principle from cognitive psychology suggests that, in a different sense, agents do not fully control their behavior. This is the principle of least cognitive effort – the principle that 'we prefer to engage less rather than more effort to obtain the same reward' (Inzlicht and Campbell 2022, 1035). This principle explains a wide range of results regarding deliberation and choice (Kool and Botvinick 2018). We argue that understanding the subtle and silent ways that this principle works within the human cognitive economy suggests an interesting tension in how agents form plans and guide their behavior. In short, behavior that seems, to the agent, to be guided and directed by the agent, may often be driven by a subtle, unnoticed kind of effort avoidance.

This can happen across a range of different kinds of cases. One case in point, upon which we focus, is that of things done by habit. Habitual actions are typically performed unthinkingly, in response to the presence of certain environmental cues that trigger a series of bodily movements on our part. For example, we often end up scrolling the news feed of social media out of habit, just because we happen to glance at our phone, and our finger movements are triggered by this environmental cue (see Neal and Wood 2009). More generally, things done by habit involve little cognitive effort insofar as they do not require any cognitive control (if anything, we need the latter to override our habits).

Now, Douskos (2017) points out that many habitual actions may be automatic in two senses: both in the phenomenological one of involving little awareness, and in the causal / explanatory one of not involving intentions or other causally relevant mental states. The former does not imply the latter (Fridland 2017; Brozzo 2021), and indeed not all habitual doings involve both. Indeed, there are a number of habits that we intentionally acquire, with good reason (Bratman 1989; Douskos 2017; Neal and Wood 2009). Even so, we will argue that habitual actions often appear attractive within deliberation precisely because they represent a low effort option. Agents may thus wind up acting through a sort of behavioral rut, simply because of the ease of doing so.

We are also often not fully in control of the way in which we fail to do some things, and do others instead. For example, some of our intentions to do something get endlessly postponed. A notable case of when this happens is one in which our intentions specify an action that is novel to us – for example, that of creating a work-related website for the first time.

Such cases are very pervasive in our ordinary lives. We will argue that such cases suggest that much behavior is not fully explained by the agent's relevant intentions or goals – in addition, the agent's disposition towards the expenditure of effort plays a subtle role in their planning. This disposition is often, though not always, in accordance with the principle of least cognitive effort. Because of this, agents often end up doing what involves the least cognitive effort. In the case of habit, this is because the link between environmental cue and bodily movement takes care of triggering our bodily movements. More cognitive effort would be involved in overriding this link. In the case of nudges (Thaler and Sunstein 2021), more cognitive effort would be involved in overriding the nudges provided by environmental cues. In the case of novel actions being endlessly postponed, a possible explanation afforded by the principle of least cognitive effort is

that novel actions are unfamiliar, and, because of this, often ones for which we lack a sufficiently detailed representation of how they should be performed. Because of this, rather than facing the cognitive effort that will be involved in figuring out how to do them, we end up doing something different. (It is a further deep and difficult question whether this avoidance is mediated by feelings rather than by a cost-benefit calculation—see Bermúdez and Massin 2023).

In short, a lot of what we end up doing (such as scrolling the news feed of social media) or not doing (such as setting up a work-related website for the first time), is not, or not entirely, under our control, and seems better explained by the principle of least cognitive effort. The principle of least cognitive effort works subtly, and silently. Subtly, because cognitive effort shows up in deliberation as a negative coloring on certain action options. Silently, because we often fail to recognize that effort was a major factor in our decision. Lower effort as represented by an action option may not accord with the amount of effort actually required.

In pointing out the principle of least cognitive effort, we do not mean to spell doom for the possibility of a more pervasive amount of control over our doings. Indeed, it is possible to become aware of the ways in which what we do is shaped by the principle of least effort, whether in the form of environmental cues and habitual links to them, or other forms of cognitive shortcuts available to us. We can, indeed, hijack these mechanisms in order to serve our goals (Duhigg 2012). Control is within our reach. It just happens not to be the norm in the way that standard philosophy of action would have us believe.

Çelik, Bartuğ (Central European University) and Agnes Melinda Kovacs (Central European University)

Communication induced belief attribution in infancy: an unexpected altercentric bias in eighteen-month-olds' pointing behavior

The current study investigates whether infants can represent that others' beliefs may be induced via communication. Significant parts of our knowledge and beliefs are formed through communication and social learning, and even infants show a strong trust in such information (Mascaro & Kovacs, 2022). However, while infants also seem to track others' beliefs based on what they have had visual access to (Baillargeon & Scott., 2017 but see Kulke et al., 2018; Dörrenberg et al., 2018), surprisingly, research targeting whether infants understand that others' beliefs can be induced via communication is scarce. Earlier research suggests that once infants have attributed a belief based on what an agent has seen, they can use third-party communication to update their own earlier belief attributions (Song et al., 2008; Tauzin & Gergely, 2018), or even try to update the agent's beliefs via communicative pointing (Knudsen & Liszkowski, 2012). However, whether they can attribute beliefs solely triggered by communication, where they cannot rely on visual access, to our knowledge, was not previously investigated. To this end, we relied on a paradigm developed by Knudsen and Liszkowski (2012) which found that 18- and 24-month-old infants corrected others' false beliefs by pointing in a communicative context. However, in that study, the agent's belief was computed relying on direct observation of what an agent has seen, and not communication. Specifically, after the agent herself placed an object in a certain location, the location of the object was changed by a confederate with the agent either being absent (False Belief, -FB condition) or present (True Belief, -TB). When the agent came back, infants pointed more to the current location of the object in the FB than in the TB condition, which was taken as evidence for i) tracking the agent's false belief and also for ii) modulating their own communicative behavior in order to correct this false belief.

In the present study, we aimed to investigate i) and ii) in a context in which a belief is attributed based on communication. In particular, 34 eighteen-month-old infants were presented with a scenario in which an experimenter invisibly hid a toy in one of three cups. Neither the participants nor the agent, sitting next to the participants, could see where the toy was hidden due to an occluder. After the removal of the occluder, the experimenter informed the participant and the agent about the location of the toy via verbal communication ("The toy is here") and pointing. Next, the experimenter changed the toy's location from one cup to the other, while the agent was in the room watching the change (TB condition) or she was not

in the room (FB condition). After the agent came back to the room, she approached the cups at the midline and stopped briefly in two different pre-determined spots, and asked the participant "Shall we continue playing?", following Knudsen and Liszkowski (2012). This phase was the "spontaneous-response phase" in which we measured whether infants pointed to one of the cups.

In FB condition, if no pointing occurred to the current location of the toy in the spontaneous-response phase, the agent checked the empty cup. Then, the agent asked the participant a direct question about the location of the object ("Where is the toy?"). In TB condition, after the spontaneous-response phase, after the participant pointed or the time elapsed, the agent found the object in its current location.

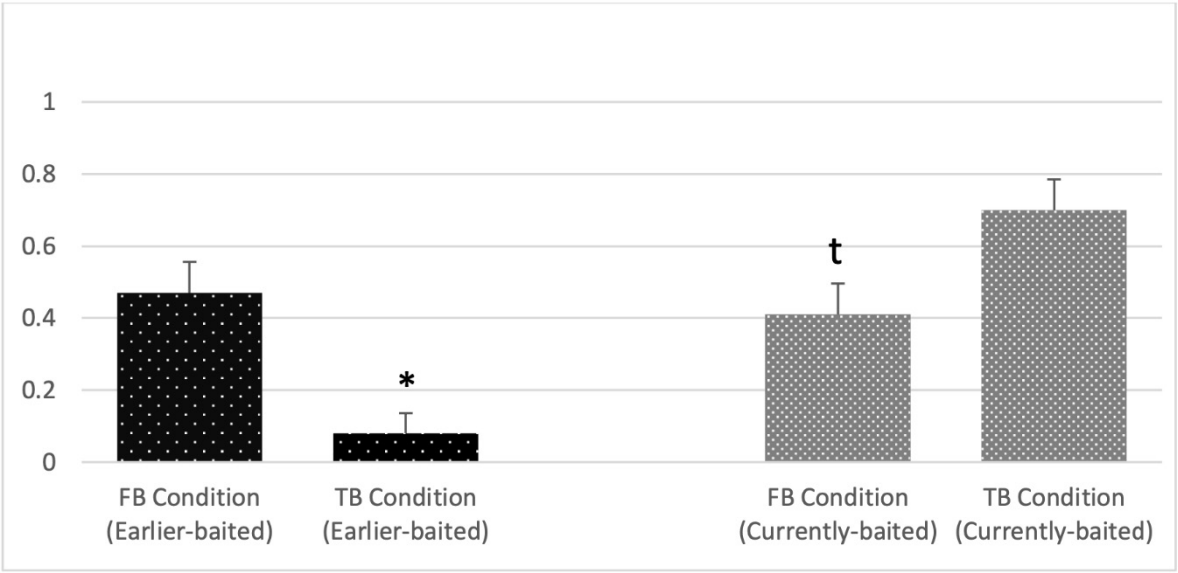
We compared infants' pointing to the currently baited location as well as their pointing to the earlier baited location (where the agent believed the object to be) in the two conditions in the "spontaneous-response phase". 18-month-olds' mean number of pointing to the currently baited location marginally differed across FB (M = .411, SD = .668) and TB conditions (M = .691, SD = .912), $t(33) = -1.926$, $p = .062$, with more pointing in the TB condition (contrary to the findings of Knudsen & Liszkowski, 2012). Interestingly, however, 18-month-olds' mean number of pointing to the earlier baited location differed across FB (M = .470, SD = .777) and TB conditions (M = .088, SD = .287), $t(33) = 2.792$, $p = .008$, with more pointing in the FB condition, reflecting an unexpected and novel altercentric effect (See Figure 1).

Furthermore, more infants (15/34) pointed to the earlier-baited cup at least once in FB condition than in TB condition (4/34), McNemar p =.009, but a similar number of infants (20/34) pointed to the currently baited cup at least once in TB condition and in FB condition (15/34), McNemar p = .267 (See Figure 2). However, when asked a direct question about the object's location in the FB condition, 74% of infants pointed correctly, Binomial p = .009.

While we expected to find evidence for i) and ii) in line with the earlier described study, we found evidence for i) only and a novel effect. Infants in our study showed different behavior in the two conditions, which suggests that they have computed the agent's FB, however, unexpectedly they have shown a strong altercentric bias and pointed more to the empty location (where the agent believed the toy to be in FB compared to TB condition).

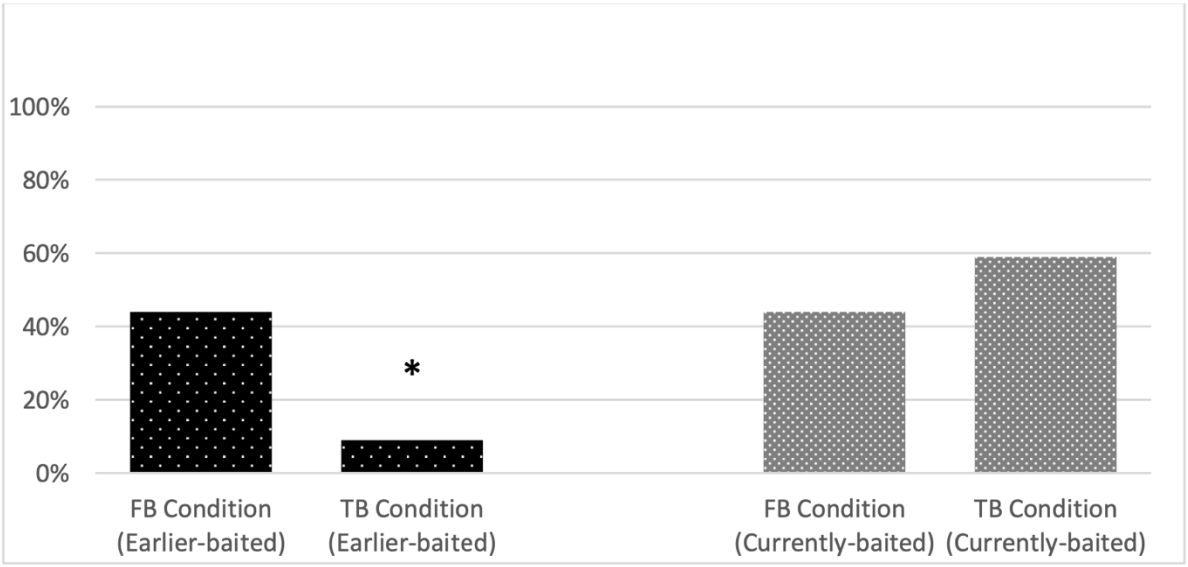
We believe that this result reflects that infants spontaneously track an agent's beliefs, computed based on communicated information and these representations seem to be so strong in this context that even their pointing behavior is influenced by the agent's false belief. Note, however, that at the same time infants could correctly answer the direct question in the FB condition, likely reflecting their ability to sustain two representations simultaneously. Whether such effects are specific to communicative contexts and the way the belief was induced and how robust they may be, are questions for future studies. In order to provide further support for this novel altercentric effect, we are currently running a replication of the study, data collection will finish in the next month and will be ready to be presented.

Figure 1.
Mean number of points per trial to the currently- and earlier-baited cup



Asteriks indicates a significant difference and t represent a marginal difference. Error bars represent standard error.

Figure 2.
Percentage of infants who pointed at least once to the currently- and earlier-baited cup



Asteriks indicate a significant difference.

References

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in cognitive sciences*, *14*(3), 110-118.

Knudsen, B., & Liszkowski, U. (2012). Eighteen-and 24-month-old infants correct others in anticipation of action mistakes. *Developmental science*, *15*(1), 113-122.

Mascaro, O., & Kovács, Á. M. (2022). The origins of trust: Humans' reliance on communicative cues supersedes firsthand experience during the second year of life. *Developmental Science*, *25*(4), e13223.

Song, H. J., & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental psychology*, 44(6), 1789.

Tauzin, T., & Gergely, G. (2018). Communicative mind-reading in preverbal infants. *Scientific reports*, 8(1), 1-9.

Coudray, Quentin (Institut Jean Nicod (France) / University of Milan (Italy))

"Appearance Proto-Concepts" as The Upper Limit of the Admissible Contents of Vision

Philosophers of perception strongly disagree about what properties one can literally see, or what properties are the "admissible contents" of vision. There remains a strident controversy as to whether we can see objects as pertaining to abstract kinds, or whether we can only see their concrete, sensory properties. "Liberal" theories accept that one can see an object as a pen (functional kind), as an animal (natural kind), or as Joe Biden (identity property). In opposition, "conservative" theories argue that we can only see an object's concrete properties, such as a pen's specific shape, colors, and texture, but we cannot see it as a pen. This is thus an epistemic sense of the word "see" that the admissible contents of vision debate is focused on. If I represent an object in front of me as a pen but it turns out that it is a toy laser, the conservative will claim this is only a misjudgment, whereas the liberal can claim that this is literally a case of misperception (i.e., a visual illusion).

I argue that neither the conservative nor the liberal positions as currently existing in the debate are satisfactory. I propose a "moderate" alternative, pursuing and empirically updating ideas previously expressed by Jack Lyons (Lyons, 2005). I reject the conservative idea that we can only see concrete low-level properties, but at the same time refuse to fly as high as liberals in claiming that we can perceive abstract kind properties such as functional kinds, natural kinds, or identity properties.

To defend my moderate view, I leave aside methods based on phenomenal introspection (Siegel, 2011; Siewert, 1998), and instead consider recent proposals in the philosophical literature that leverage empirical data to defend perceptual Liberalism (Bayne, 2009, 2016; Bayne & McClelland, 2019; Block, 2014, 2023; Butterfill, 2009; Calzavarini & Voltolini, 2022; Fish, 2013; Ransom, 2020a, 2020b; Stokes, 2018, 2021). A central argument upheld in these publications is to point out that empirical results in vision science show that categorizing objects into high-level abstract kinds often bears typical behavioral markers of vision: it can be extremely fast, be subtly stimulus-dependent, be automatic and ballistic, operate outside of attentional focus, and yield adaptation aftereffects. A simple and straightforward explanation for such behavioral markers is that categorizing objects into high-level kinds can be done in vision, and does not require to marshal cognitive processes, as perceptual conservatives wrongly predict.

I agree with liberals that such "common marker" arguments give us good prima facie reason to reject conservatism. Nonetheless, there is an important problem arising.

One empirically plausible mechanism explaining how vision is capable of categorizing objects into high-level kinds is that of appearance similarity comparison. For instance, our visual system can compute that a thin, straight, pointy cylinder is similar to a particular class of previously encountered object representations stored in visual long-term memory that are also thin, straight, pointy metallic cylinders. That visual categorization is similarity-based is empirically backed up by a long tradition in cognitive psychology (Goldstone, 1994; Hahn & Ramscar, 2001; Medin et al., 1993; Nosofsky, 1986; Posner & Keele, 1968; Reed, 1972; Sloman & Rips, 1998; Sloutsky et al., 2007; Tversky, 1977). One could even claim that similarity-based categorization is the only kind of mechanism that could ground visual categorization, since a rule-based mechanism of categorization seems to require cognitive processes for the manipulation of rules and inferences.

But this is where a problem arises: a conservative could claim similarity-based categorizations in vision demonstrate that one does not visually represent abstract kinds after all, but only meaningless classes of similarity. The abstract functional property of "being a pen" is different from the mere similarity class of thin, straight, pointy metallic cylinders: one could perceive this similarity class without perceiving pens (the functional kind). This would be enough to successfully complete categorization tasks, thus providing the conservative an explanation for why subjects seem to recognize abstract properties (pens, animals etc.) in vision: they simply track appearance similarity classes coextensive with these properties. The common marker argument seems insufficient.

Nonetheless, I do not think visual conservatism is vindicated. This is because visual representation of appearance similarity is not comparable to a low-level representation. In fact, as philosophers have noted (Goodman, 1972) and as empirical research confirms (Dubova & Goldstone, 2022; Goldstone et al., 2001), similarity relations are not determined by low-level properties in any necessary way. Rather, similarity relations are built by the visual system to fulfil specific tasks in specific contexts. In a nutshell, similarity is meaningful: it wouldn't be visually represented if it was not coextensive with classes of properties that are behaviorally relevant for the perceiver. One does represent the class of similarly thin, straight, pointy metallic cylinders, the class of "pen-appearances", because pean-appearances are in this world reliably coextensive with the presence of pens.

Even if it turns out that a pen-like object is not really a pen, but rather a laser toy, a pen-appearance representation would still be accurate. Inaccurate appearance representations can occur when for instance the stimulus is ambiguous, and when similarity computation is thus made challenging. Inaccuracy can also emerge when one uses a task or context-irrelevant dimension of similarity comparison, as when a novice birdwatcher trying to discriminate between greenfinches and sparrows uses beak color as a similarity comparison dimension, while differences in beak color are not coextensive with the distinction between greenfinches and sparrows. One here computes the wrong kind of similarity.

I conclude that visual "appearance kind" representations based on similarity comparisons form the upper limit of what can be visually represented. As they are semantically in-between the concrete low-level representations of conservatism and the abstract high-level representations of liberalism, we should embrace a kind of moderationism in the admissible contents of vision debate.

Cutting, Nicola (York St John University) and Darcy Neilson (York St John University)

Capturing children's innovations in different contexts; comparing structured problems and play

Whilst many species make and use their own tools, it is only humans who have developed an abundance of complex tools that we use in almost every aspect of our lives. Two factors are thought to drive the development of our tool-rich world: faithful imitation, i.e., watching and copying how others make tools, and innovations or modifications (Legare & Nielsen, 2015). Most research has focused on social learning and has indeed demonstrated that humans are extremely faithful copiers (see Hoehl et al., 2019). There has been a smaller focus on innovations.

Work by myself and colleagues was the first to investigate children's capacity for tool-innovation. We used a problem-solving paradigm that tasked children with fishing a bucket out of a tube to retrieve a reward contained inside (Beck et al., 2011). Children were given materials such as pipecleaners and string to solve the task, with the solution being to bend a pipecleaner into a hook. Children found this task extremely difficult with children aged 5 rarely innovating a hook tool and only half of children succeeding by age 8. This finding has been replicated by several research teams across the world (Nielsen et al., 2014; Voigt et al., 2019) and has been shown to be a stable finding in western and non-western populations (Neldner et al., 2017).

Whilst the controlled structure of these studies allows researchers to explore the mechanisms underlying children's abilities, the format of these experimental studies may not allow us to capture the true innovative abilities of young children. Indeed, children's lack of innovation and creativity in these tasks appears at odds with the imaginative, creative play observed in children. In artificial task environments, children are tasked with working individually on very structured problems, with short time limits imposed. Real-world innovations are unlikely to unfold in this way. They are likely to take time, be the product of collaboration and occur naturally rather than in a test situation.

To investigate how context may play a role in children's ability to innovate we designed a study to compare capacity for innovation and measure creative behaviours in structured and non-structured environments. The study presented measures innovation during a play context and compares it to innovation in a more structured, task environment. Children aged 4 to 7 (N=120) participated in pairs to facilitate play behaviour and collaboration. Pairs were presented with two sets of apparatus/materials. Each set of apparatus was designed to afford innovation but was also an attractive play scenario for the children. The first was pirate themed and consisted of a pirate island with treasure hidden at the bottom of a long tube, and a pirate ship with pirate characters and pipecleaners (see Figure 1.). This scenario was based on the hooks paradigm (Beck et al., 2011), and children could retrieve the treasure by making a pipecleaner hook. The second scenario was alien themed and consisted of an alien planet scene with a small egg-shaped object at the bottom of a long tube (See Figure 2.). Presented alongside this was a jug of water. This scenario was based on the floating peanut task (Hanus et al., 2011), and children could pour water from the jug into the tube to raise and retrieve the egg object. Children received the two sets of apparatus in counterbalanced order, and each was counterbalanced as to whether it was presented in a play context, e.g., 'Here are some things you can play with', or as a structured task, e.g., 'Can you help the pirates retrieve the treasure?'. Pairs were given 10 minutes to interact with each set of apparatus.

Findings from the study indicate that children were extremely problem focused, and once the treasure or egg was discovered, persisted in trying to solve the problem (i.e., retrieve the object) in both the structured task condition and the play condition, even when they were not tasked with solving the problem in the latter condition. This resulted in similar innovation rates and no difference between the play and structured conditions for the pirate scenario ($p = .630$) or the alien scenario ($p = .176$). Ongoing analysis is investigating whether there are differences in children's creative and exploratory behaviours between the two conditions (Play vs. Structure) and between the two scenarios (Pirates Hooks vs. Alien Floating toy). Initial coding using the Analysing Children's Creative Thinking (ACCT) Framework (Robson & Rowe, 2012) points towards children in the play condition engaging in more exploratory behaviours and exhibiting higher involvement with the apparatus. Full analysis of these behaviours will be presented and discussed, with a focus on whether and how we can design rigorous empirical studies that capture the true extent of children's capacity for innovation.

Dahl, Niklas (Lund University)

Meaningless Communication: Towards an Understanding First Approach to Language

In most cases when we're talking to someone we'd like to be understood. But despite how central it seems to our communication, the notion of understanding itself hasn't received a lot of attention. What does it mean to understand an expression or an utterance? Usually, this question is approached through the concept of meaning; understanding an expression is merely having the right relation to its meaning.

What I want to do here is to explore what happens when we don't accept this order of explanation. There are, as I will argue, some reasons to think that this meaning-first approach to language makes it difficult to explain how we can have the knowledge we need in order to understand each other. While hardly conclusive, I take it that these considerations motivate an independent look at what it means to understand.

I will begin by looking at a parallel discussion about the epistemic concept of understanding. From there, I will take the central idea of understanding as a kind of skill or knowledge-how. This skill has, I will argue, both a practical component of knowing how to use the expression correctly and a cognitive component of having a mental conception of that practice. Thus, understanding an expression will be like possessing the twin skills of using it in conversation and reason. Thus, understanding at the level of an expression, is quite literally taken to be a kind of competence. From there, it's a matter of giving an account of understanding an utterance; comprehending a token of an expression. In slogan form, comprehension is conversationally deployed competence.

Next, I will make the point that a single expression can simultaneously fulfil several linguistic functions. Since these functions are fulfilled through different underlying practices, I will argue that we have to relativise understanding of an expression with respect to the different functional roles it can play. That is, there are several ways to understand the same expression depending on what roles it can play. I will borrow the notion of linguistic function from the framework of systemic functional linguistics and argue that to properly analyse what it takes to understand a particular type of expression we first need to identify what different communicative functions it fulfils. As an example, singular terms play at least two different roles in communication: a talking-about function and a picking-out function. The first enables interpersonal exchanges of information by providing a public tag, whereas the latter pertains to the use of singular terms to co-ordinate action. As these functions are underwritten by distinct practices of use, we require distinct skills to understand a term with respect to the two. In this way, we can see that from the understanding-first point of view there are two notions of reference; two ways which a term can be about something. And, as such, two distinct ways to understand a singular term.

The last part of my account of understanding is going from expressions to utterances. Essentially, the idea is that this is a matter of deploying the skill of understanding the expression to the particular conversation where its uttered. As I will argue, this is a kind of active engagement with both our own prior competence with the expression and that of our conversational partner. That is, understanding what someone says is more like dancing than decoding the meaning of a message. Although I may be skilled at, say, waltz, that's not enough for me to successfully dance with you. That knowledge-how must be supported by perceptual and contextual information about your movements and waltzing ability. Similarly, when you make an utterance, I understand it by using perceptual and contextual cues about the conversational environment to deploy my general knowledge-how to use the expression you uttered a token of.

Finally, I will sketch a picture of communication based on my account of understanding. Communication, I will argue, is deemed successful when the speaker and hearer agree that they both understand the utterance. In doing so I hope to illustrate some of what I take to be the advantages of an understanding-first approach.

Dolega, Krzysztof (Université Libre de Bruxelles)

Hidden in plain sight: on valence and the elusive nature of affective perception

A growing body of work within psychology (Lebrecht et al. 2010, 2012; Barbot & Carrasco 2018) suggests that perceptual experiences can not only produce emotional responses or evaluative judgements, but are themselves affectively colored or valanced. While the debate about how pervasive this phenomenon exactly is, it seems clear that (at least some of) our perceptual experiences are not neutral, but are laden with varying degrees (and perhaps kinds) of value. This, in turn, presents the obvious question of how to conceptualize and explain this kind of affective perception (de Vignemont, 2021).

Most authors try to capture the affective aspect of perceptual phenomenology in intentional terms, but disagree whether differences in perceptual valence result from differences in the contents of perceptual experiences or differences in the attitudes directed at those contents. Furthermore, proponents of the content view disagree whether the affective character of experiences depends on representations with evaluative content (Bain 2013; Carruthers 2018) or representations with imperative content (Klein 2015; Barlassina & Hayward 2019; Martínez 2022). The main objection to the content proposal is that two subjects can accurately represent the same object yet have opposing experiences in regard to its valence. The same food item that looks tasty and enticing to me can seem disgusting and off-putting to someone else. This, in turn, suggests that what distinguishes the two experiences is not a difference in what is perceived, but in how it is perceived. The competing attitude view tries to account for such differences by postulating that affectively laden perceptual experiences are a result of sensory contents being bound up or fused with evaluative attitudes (de Vignemont 2021; Jacobson 2021). However, this view is not free from controversy, as its proponents claim that perceptual valence is a non-sensory and non-cognitive modality of its own kind. According to de Vignemont (2021, 12), the evaluative attitudes involved in affective perception are distinct from the contents of sensory processing, yet become inseparable from them once they become bound in perceptual experiences. In her own words: "once you have put on affective mental paint, you cannot go back to the [neutral] image" (2021, 13).

In this talk, I aim to create a different picture of "affective mental paint" and defend a more radical version of the attitude approach to perceptual valence in which the affective character of experiences is identified with the causal-functional properties of the vehicles of perceptual representations, and not with additional attitudes layered onto such representations. In other words, the claim defended in this paper will be that the affective valuation of perceptual experience is a result of modifications in the profiles of the vehicles which carry the (otherwise neutral) perceptual contents. I present several arguments in favor of this view.

The first line of my argumentation is neurobiological. De Vignemnont tailors her position to match the evidence for a "quick and dirty" subcortical forward connection between the visual cortex and the amygdala (2011, 3). Indeed, there is mounting evidence that the amygdala responds differently to negatively valenced stimuli at time scales which are much faster than allowed by cortical connections (Garrido et al., 2012; McFadyen et al., 2017). However, there is little evidence for a matching recurrent subcortical connection, meaning that any evaluative processing done in that amygdala must be communicated either directly via cortical pathways to the visual processing areas or indirectly through connections to the prefrontal areas responsible for attentional modulation. The latter hypothesis is more consistent with Barbot & Carrasco's 2018 study, which has demonstrated attentional fine-tuning of perceptual representations. This seems inconsistent with de Vignemont's view, as it does not require any layering of additional attitudes on perceptual contents, but only fine-tuning of perceptual states through attentional modulation. In light of this, the former hypothesis about direct modulation of visual areas might appear more consistent with de Vignemont's proposal. However, earlier functional imaging findings (Lebrecht, 2005) located neural correlates of perceptual valence in regions associated with the processing of object features, thus indicating that the direct signals from the amygdala are likely also involved in tuning perceptual representations, rather than with merely tagging the outputs of perceptual processing with evaluative attitudes.

My second line of argumentation will appeal to phenomenology as well as explanatory parsimony. According to de Vignemont: "one can no longer retrieve the 'pure' visual experience, untainted by affective coloring, once it is bound with evaluative feeling [...] seeing negatively is more than seeing and disfavoring"

(2021, 13). Indeed, but if this is the case, then the radical vehicular view defended here is much closer to this specification than the layered attitude model. The vehicular view claims precisely that it is the way in which content is represented that is tuned by recurrent processing. In other words, the view is supposed to be an account of seeing negatively, rather than seeing plus being attracted or disfavoring the perceptual content. Finally, the vehicular view is more parsimonious as de Vignemont's layered attitude view. Not only does the view postulate a simpler model of affective perception, but the way in which it answers the question about what makes perceptual experiences valenced fits well with prominent proposals about the nature of phenomenal properties (Block 1996; Papineau 2014) and the function of conscious experience more generally (Cleeremans & Tallon-Baudry 2022).

Dranseika, Vilius (Jagiellonian University)

Dreaming feels like seeing but seeing does not feel like dreaming. Folk beliefs about phenomenological differences and similarities between kinds of mental states

The topic of this paper is pre-theoretical beliefs about phenomenological similarities and differences between mental states. I report the results of a set of six studies on folk beliefs about phenomenological differences and similarities between dreaming, remembering, perceiving, imagining, and hallucinating. On the one hand, I investigate folk beliefs about how similar are experiences of being in one kind of mental state to those of being in another kind of mental state. On the other, I investigate folk beliefs about how likely it is to mistake being in one kind of mental state for being in another kind of mental state. This is also an attempt to develop ways to study folk beliefs about phenomenological differences and similarities between mental states.

First, using two different methods, I collected judgements on phenomenological (dis)similarity between mental states (Studies 1 and 2). For instance, in Study 1, I explained to the participants that the study is about what one experiences while having various mental states, about what it feels like to have various mental states, and then I gave participants triples of mental states and asked them – for each triple – to indicate which mental state feels the most dissimilar to the other two. I treated the proportion of 'wins' within a triple as a measure of phenomenological similarity. Results of Study 1 are presented in Figure 1 (see pdf version attached). In these studies, participants were inclined to treat some pairs of mental states as phenomenologically relatively similar (dreaming and imagining; hallucinating and dreaming) and other pairs as phenomenologically relatively dissimilar (hallucinating and remembering; dreaming and seeing).

Similarity judgments, however, are often sensitive to the order of comparison (Studies 2 and 4). For instance, study participants were much more inclined to agree that hallucinating feels like seeing than that seeing feels like hallucinating, and much more inclined to agree that dreaming feels like seeing than the other way around. Despite differences in methodologies, multidimensional scaling of (dis)similarity data from Studies 1-4 revealed a recurring two-dimensional structure of (dis)similarity judgments.

Finally, I asked study participants to indicate whether they find some kinds of metacognitive errors (mistaking being in one kind of mental state for being in another kind of mental state) especially likely. The general idea is that believing that two states are phenomenologically similar may be associated with believing that such states can be mistaken for one another. Importantly, however, several interesting asymmetries in similarity judgments were observed in Studies 3 and 4. For instance, study participants were much more inclined to agree that dreaming feels like seeing than that seeing feels like dreaming. Consequently, we can expect that some kinds of mistakes will be seen as more likely. Thus, we can expect that study participants will see mistaking dreaming for seeing as more likely than mistaking seeing for dreaming. Study 5 indeed provided evidence that beliefs about phenomenological similarity are associated with beliefs about how likely various metacognitive mistakes are – the more study participants were inclined to agree that X feels like Y, the more they were inclined to agree that mistaking X for Y is likely. Study 7 links these patterns in beliefs about likelihood of metacognitive mistakes with beliefs about metacognitive transparency of

various mental states – a mental state that is believed to be less metacognitively transparent than some other mental state is also taken to be more likely to be mistaken for the other state than the other way around.

Studying folk beliefs about phenomenological differences and similarities between kinds of mental states is important not only because it contributes to our better understanding of folk psychology, but it can also be useful in trying to understand psychological underpinnings of philosophical concepts and theories. For example, the fact that study participants are overwhelmingly more inclined to think that dreaming feels like seeing than that seeing feels like dreaming (and, consequently, take one kind of metacognitive error (mistaking dreaming for seeing) to be more likely than the other kind (mistaking seeing for dreaming)) could be useful in thinking about psychological appeal of dream skepticism.

Drożdżowicz, Anna (Inland Norway University of Applied Sciences)

Speaker impressions and risky inferences

A speaker may seem wise or funny in virtue of what they say. A speaker may seem anxious or sloppy in virtue of how they say it. Spoken linguistic communication is not just a channel for sharing information intended by speakers, but also a channel of unintended information about speakers. A speaker's voice and their manner of speaking provide rich information about their attitude and emotional state (Bänzigier et al. 2014), but also extra-linguistic cues that can point to various speaker characteristics, including their: age (Mulac et al. 1996), sex (Owren et al. 2007), identity (Baumann & Belin 2010), ethnic background (Wolfram & Thomas, 2008).

Impressions associated with speakers' voices and their manner of speaking are socially important and can greatly influence linguistic and social interactions for instance by influencing how hearers will evaluate speakers (e.g. Lev-Ari & Keysar 2010, Klofstad et al. 2012; Mottram 2016; Rakić 2019). It is therefore important to address the question of whether and to what extent audition-based impressions are a good source of information about the speaker and their characteristics. The goal of this talk is to make progress on addressing this question.

I will suggest that one promising way to identify especially epistemically risky inferences from audition-based impressions of a speaker is by looking at two features of such impressions: 1) the extent to which they are stimulus-dependent; and 2) the extent to which they result from exposure-contingent generalizations about various speaker characteristics. The latter feature will primarily concern impressions of socially constructed characteristics of the speaker arising from specific context-based exposure patterns and based on specific stereotypes.

Dunin-Kozicka, Monika (Catholic University of Lublin), Magdalena Szubielska (Catholic University of Lublin), Tom McClelland (University of Cambridge) and Paweł Fortuna (Catholic University of Lublin)

Affording Imagination and Virtual Reality

Tom McClelland (2020) postulates the existence of the so-called mental affordances. Affordances – standardly understood as action possibilities offered to us by various environments, objects, or situations – have been so far mainly associated with bodily actions (e.g. Gibson, 1979; Costall, 2012; Turvey, 1992). For example, a chair would afford sitting on it, an incoming ball would afford catching it, and a teacup would enable one to grip it. According to McClelland's proposal, different situations and things in the world can also invite mental actions such as paying attention to something, counting or imagining. Granting that two

basic requirements that must be met for affordances to occur is a perceptual requirement (S perceives x as affording φ-ing) and a potentiation requirement (S perceiving x as affording φ-ing potentiates S φ-ing), McClelland argues that both of them would also apply to mental affordances, including imaginative affordances, which are the leading subject of our study.

Our main research question is how immersion in artistic virtual reality (VR) affects the perception of imaginative affordances, or — to be more specific — what impact such immersion has on one's perceiving the possibilities of imaginative actions in real-world objects. When dealing with imaginative actions, we identify them not only with mental visualizations or quasi-perceptual experiences, but also with propositional imaginings in the form of counterfactual thoughts or suppositions (see Kind, 2022a; Liao and Gendler, 2019). We assume that VR immersion in the art world of this type may result both in seeing more imaginative affordances in the real-world objects, and noticing new such affordances.

In the study, we use the original VR simulation of an art showcase, i.e. the Museum of Other Realities (https://www.museumor.com), which contains over 40 virtual artistic works made by different authors. The Museum allows visitors not only to experience unusual works of art (e.g. those that fluently and responsively change their shapes, colors, location and other perceived properties), but also to experience unusual forms of interactions with them (such as, i.e., immediately changing the perspective of seeing the works of art; "entering" them; transforming them by the use of one's voice). Based on our analysis and own experiences with VR, we argue that immersion in such a world is likely to make one more sensitive to those characteristics of physical objects that may be subject to various imaginary alterations, and thus enable one to perform more imaginative actions with these objects as well as to perform qualitatively different or original imaginative actions. In addition — granting the possibility of exercising and improving people's imaginative skills (Kind, 2022b, 2020) — we can expect that some of the virtual simulations can be successfully used as tools for imagination training.

Our experiment is designed so as to test all these expectations. First, the participants from the experimental group experience a 20-minute immersion in the Museum of Other Realities, while the subjects from the control group watch a 20-minute documentary about an art museum. Next, the subjects are presented with an everyday object (e.g. fork, box, tape, etc.) and are asked to answer the following questions in turn:

1. What do you think about when you look at this object?
2. What can you do with this object? List as many possibilities as you can think of.
3. How could this object be used in an art gallery?
4. What can you imagine when you look at this object? List as many possibilities as you can think of.

Additionally, the subjects fill in the scales measuring the controlled variables (i.e. interest in art, VR experience and immersion capabilities).

Our research takes into account the recent philosophical concept of imaginative affordances (McClelland, 2020, 2021; McClelland & Dunin-Kozicka, in preparation), and combines it with the use of the latest immersive technology. It not only allows us to verify the exploratory and explanatory capabilities of the concept of mental affordances, but also lets us join the recently important philosophical and psychological discussion on imaginative skills and the possibility of their training by some viable means (Kind, 2022b, 2020). We expect that the proposed combination of philosophical and psychological methodology will allow us to formulate an important voice in the social discourse on imagination and virtual reality.

The experiment presented above is an ongoing study, scheduled for completion in June 2023, with final statistical and theoretical analysis expected by July 2023.

Ebel, Sonja J. (Max Planck Institute for Evolutionary Anthropology; University of Leipzig), Kathrin S. Kopp (Max Planck Institute for Evolutionary Anthropology; University of Leipzig), Evgeniya Kirilina (Max Planck Institute for Human Cognitive and Brain Sciences), Ilona Lipp (Max Planck Institute for Human Cognitive and Brain Sciences), Heather Cohen (Max Planck Institute for Evolutionary Anthropology), Catherine Crockford (Max Planck Institute for Evolutionary Anthropology; Institut des Sciences Cognitives, CNRS), Daniel B.

M. Haun (Max Planck Institute for Evolutionary Anthropology) and Roman M. Wittig (Max Planck Institute for Evolutionary Anthropology; Institut des Sciences Cognitives, CNRS)

Developmental trajectories of tool use in zoo and sanctuary-housed chimpanzees

Chimpanzees are one of our closest living relatives. In addition to researching chimpanzees to understand their biology and nature, studying them comparatively can provide new insights into the cognitive makeup of our last common ancestor. We know that humans have a slow life history with extended developmental trajectories, which includes a period of learning how to make and use tools. Chimpanzees also demonstrate slow life-history patterns and a wide variety of tool use. However, it is still unknown how tool-use abilities in chimpanzees develop, particularly when chimpanzees learn about the functional properties of the tools used. To explore these developmental trajectories, we conducted a tool-use experiment with 77 zoo and sanctuary-housed chimpanzees, aged 1 to 53 years. The test setup mirrored everyday situations in which chimpanzees use sticks as tools to access food sources. For this purpose, we provided tubes from which juice could be extracted by dipping, alongside standardized sets of sticks varying in properties related to their tool functionality: long vs. short, thick vs. thin, solid vs. frayed. We looked into the developmental trajectories in terms of tool choice, i.e. functionality and efficiency, as well as precision and speed in using the tool. The results are discussed with regard to their significance for comparative research and their implications for the origins of human tool use.

Erden, Y J (University of Twente)

AI for psychiatry: close encounters of the algorithmic kind

This paper explores the use of artificial intelligence (AI) in psychiatry and argues that AI development too often overlooks the complexity and necessary imprecision of the theories it adapts from fields like psychology. In psychiatric contexts, AI can spot patterns and generate predictions, e.g. 'big data' analysis via statistical learning-based models [1]. In these ways, AI offers scope to automate some apparently routine steps in psychiatry, and could help to improve efficiency, mitigate clinician bias, and offer predictive potential, including through the analysis of neuroscientific data [2,3]. Electroencephalography (EEG), for instance, promises data on brain activity related to cognition, plus emotions and behaviour [4]. Such brain data offers apparently objective accounts of what otherwise requires interpersonal engagement and observation.

Yet theories in psychology are not neutral, and any biases in such theories or approaches [5,6] can (and do) find their way into AI applications [7]. This in turn helps to encode and reify such values and judgements. In addition, AI systems can (and do) perpetuate accounts of human minds, brains, and behaviours that are based on shaky foundations [8], or studies that have been retracted or discredited. There is now substantial evidence that well-known studies in psychology do not hold up in replication studies. Reasons for these problems include bias (whether unintentional, methodological, confirmation, negative etc.) and fraud (methods, reporting of results etc) [9,10]. A seminal study in psychology in 2015 demonstrated the shocking extent to which there had been a failure to replicate, with reasons that span a lack of funding or commitment to the principles of open science, plus limited value ascribed to the testing and retesting of scientific results, whereby too few studies verify 'well-established' scientific results [11] (e.g. novelty is prioritised over replication). Some are yet to face such scrutiny [12].

Even where research in psychology is sound, there is more to psychiatry than can be automated, yet these elements are sometimes deprioritised in the face of an increasing mental health crisis [13]. AI analysis of big data for predictive purposes could eventually supplant the phenomenological perspectives that underlie a person's actions, choices, and experiences, as well as bypass the necessarily discursive engagement between patient and clinician. Brain data can improve explanatory models, but this should not be at the expense of essential qualitative practices. Technological methods for assessment and diagnosis

might seem time and cost efficient, but there remains an important role for (even imperfect) interpersonal methods in medicine and care. Psychiatry includes the assessment and diagnosis of illness and disorder within a largely interpersonal communicative structure typically involving a physician and a patient, and AI should not disrupt this.

Some core principles for the use of AI in psychiatry are thereby proposed, including recognition that:

i. AI lacks capacities crucial to psychiatry, e.g. nuance / flexible thinking about mental states (beliefs, desires, intentions, preferences, needs), plus context / experience. The use of AI in psychiatry should not undermine the necessary relational aspects of care, especially where technological fixes seem to offer respite for resource heavy fields.

ii. Statistical data / analysis include inferences, while models of psychiatry are located within normative frameworks [14]. AI should not cement simplistic classifications or exacerbate harmful biases, and care should be taken in the selection of theories of mind, brains, and human behaviour. AI systems need to be flexible enough to adapt as these theories likewise adapt, or to change tack where theories are discredited or papers retracted.

iii. Brain data needs particular scrutiny given potential to bypass self-reporting / interpersonal, discursive methods [15].

iv. AI should be sufficiently transparent, with methods, processes, (brain) data sets, including for training, open to critique [16].

AI can play a role in psychiatry, e.g. by making explicit seemingly intuitive inferences, but it cannot achieve this if it replicates biases and flaws, in design and datasets, or perpetuates outdated or otherwise flawed theories that are currently being deconstructed in psychology and psychiatry as a result of the replication crisis.

Ewing, Kyley (Cape Breton University) and Natasha Gallant (University of Regina)

Perceptions of Time and Freedom in Dementia and Its Effects on Well-Being

This paper applies insights from our leading philosophical theories of free will and the phenomenology of the passage of time to the data on dementia and well-being. Along the way, we also address the relationship between both situational and overall life levels of well-being and the human experience of freedom over time. There are four main parts to our analysis of the perception of the passage of time and free will to well-being.

First, we begin by introducing recent research that identifies trends in distortions of time perception found in those with dementia. Individuals with dementia often underestimate the passage of time (see Papagno, Allegra, & Cardaci, 2004 or Iwamoto & Hoshiyama, 2012) and imagine that the future is more open-ended than individuals of comparable ages without dementia (Grewal, 1995 or Shomaker, 1989). As noted by Lang and Carstensen (2002), when individuals perceive time as open-ended, they prioritize future-oriented goals which focus on broadening horizons and knowledge acquisition.

Second, we delve into the conscious processing of the passage of time within the human mind and suggest that, in line with the B-theoretic, block universe, the passage of time is purely a phenomenon of the human mind rather than part of the objective reality of our world (see Dainton 2010 or Silberstein, Stuckey, & McDevitt, 2018). It is therefore no surprise that those with disfunctions across the cognitive domains found in individuals with dementia should experience the passage of time differently than those without such distortions. Importantly, both groups would be equally connected (or disconnected) to the objective truth about temporal passage.

Third, we then defend a compatibilist account of free will (see Kane 2005) and propose that there is an important relationship between our perception of temporal passage, free will, and well-being that is highlighted by the experience of those with dementia. In particular, our suggestion will be that, regardless of chronological age, perception of our own freedom is deeply connected to our perception of the time that has passed versus the time that is still to pass in our lives.

Further, insofar as happiness is intimately connected to the choices we perceive ourselves as having and the openness we feel is offered by our own future, those with dementia should, at least in some ways, have a heightened sense of well-being when compared to those of a similar chronological age without dementia.

Finally, we conclude that, as shown by the available research with individuals who are living with dementia, the human sense of well-being has a strong basis in both our perception of time and our sense of our own free will. Although not possible in our current analysis, one avenue of future exploration would be to focus in on the various types and degrees of dementia in order to further understand the role of the perception of temporal passage and freedom in well-being.

Facchin, Marco (IUSS PAVIA)

Neural representations unobserved

Philosophers supporting the "cognitive neuroscience revolution" [1] claim that cognitive neuroscience operates with a rich and stable concept of neural representations. In their view, neural representations are map- or model-like structures that represent their targets by being structurally similar to them [2-4]. Call these representations neural structural representations (NSRs for short). NSRs have two key features. First, their representational vehicles have certain relational properties that make them observable [cf. 5,6]. Indeed, we can observe and notice the structural similarity tying them to their targets. Secondly, since such a structural similarity is both semantically relevant and a causally efficacious property determining how vehicles are processed, the semantic properties of NSRs play an active causal role within our neurocognitive mechanisms [cf. 7,8].

Notice that these two features entail that, in order to know whether our neurocognitive mechanisms really operate on NSRs, we can simply observe the swirl of our neuronal activity, and see whether it actually instantiates vehicles having the desired properties. In my talk, I will observe neuronal activity at three distinct functional scales, claiming such vehicles cannot be observed in any of these scales.

First, I will consider the scale of individual neuronal activity. Whilst some have argued that individual neurons underpin NSRs [cf. 9-11], a proper scrutiny of their activity reveals that this is not the case: the relevant structural similarity tying them to their target fails to play the required causal role within our neurocognitive system. The reason, bluntly put, is that the relevant structural similarity is grounded on the temporal coincidence holding between individual neuronal activations and target presentations [see 11]. Violations of such a temporal coincidence, such as the one caused by the "recycling" of neuronal activations via off-line simulations [e.g. 12], thus worsen the similarity. Now, were individual neuronal responses NSRs, a worsening of their similarity should result in a lowering of an agent's odds of behavioral success [2,7]. But clearly this is not the case: off-line simulations are hugely beneficial to an agent, and significantly improve their chances of success [13,14]. Hence, individual neuronal responses do not play the relevant causal role associated with NSRs, and thus fail to qualify as such.

In second place, I will consider the scale of neuronal maps, considering various such maps, ranging from the retinotopic map in V1 to the sensory and motor "homunculi" [cf. 15]. I will claim that, appearances notwithstanding, these maps do not in fact qualify as vehicles of NSRs. Cortical maps too fail to play the causal role that characterizes the vehicles of NSRs. My argument, however, will be comparatively more straightforward: in the case of vehicles of NSRs, higher degrees of structural similarity should correspond to higher chances of behavioral success [see again 2,7]. But, in the case of many cortical maps degrees of similarity and odds of behavioral success are either uncorrelated [eg. 16] or even anticorrelated [eg. 17,18]. Thus, in spite of their undeniably map-like appearance, neuronal maps fail to qualify as proper vehicles of NSRs.

Lastly, I will consider the structural similarity holding between the activation spaces of various neuronal regions and the target domains these regions respond to, as revealed by neurocomputational modeling techniques [e.g. 19] and data analysis techniques such as Representational Similarity Analysis [e.g. 20,21]. I will argue that whilst in this case we actually observe a structural similarity playing the desired causal

role, such a structural similarity holds amongst the wrong sort of things, and so it too fails to underpin any NSR. This is because the relevant similarity does not hold between any individual neural vehicle tokened in the brain and its target, but rather sets of individual vehicles and sets of targets. But sets are definitely not concrete, hence they cannot be representational vehicles, let alone vehicles of neural representations.

NSRs thus not tokened within our brain - at least, not at the three functional scales where they are most likely tokened. This, I will conclude, poses a significant challenge to the "cognitive neuroscience revolution".

Feyen, Tess (Ecole Normale Supérieure) and Alda Mari (Ecole Normale Supérieure)

The (quasi-)objectivity of moral predicates

Moral predicates (like good, wrong) gained more popularity in the last years (Ruiz 2019, Stojanovic 2019, Ruiz & Faroldi 2022), and linguists often compared them to predicates of personal taste (i.a. Lasersohn 2005) and with aesthetic predicates (i.a. Stojanovic & McNally 2014). Indeed, moral predicates show similar patterns to predicates of personal taste in that they express a subjective judgment. Faultless disagreement appears when two speakers disagree on a subjective matter (for example, taste). A statement made about taste cannot be countered by stating the speaker's experience is false.

1. Marc: Sea urchin is tasty.
Marie: No it's not!
2. Marc: Abortion is wrong.
Marie: No it's not!

If the semantics of the subjectivity expressed in PPTs have been accounted for in great details (i.a. Kennedy & Willer 2016, 2020), the subjectivity expressed by moral predicates is still in question (Stojanovic & McNally, 2022). Moral predicates (see 2.) don't seem to lead to faultless disagreement, in the sense that Marc and Marie could continue their conversations until they find a consensus. This conflict cannot be resolved for PPTs, since two speakers cannot share directly their taste buds experience with each other.

On the contrary, it seems like the expression of moral claims is more than just a description of an experience. Marc could very well think that abortions are wrong without having ever experienced it, and as such he is not describing anything about his personal experience of the world. Instead, he is making a general claim about the definition of wrongness, including abortions in this definition.

This difference between predicates of personal taste and moral predicates is also shown in their felicity conditions when embedded with attitude verbs like "find" and "consider".

3. a. I find this bottle empty.
b. I consider this bottle empty.
4. a. I find lying to be worse than stealing.
b. I consider lying to be worse than stealing.

"Find" is much more restrictive than "consider". Corpus studies have also shown that moral predicates prefer to appear with "consider" rather than with "find" (Stojanovic & McNally 2022). Kennedy & Willer (2016, 2022) have given an extensive account of the difference between "find" and "consider". Their main theory is that a subjective attitude asserts a belief in a proposition that could also be counterstated. Find-statements are defined by a specific form of subjectivity, that they call "radical counterstance contingency". Find-statements are radically contingent because they are essentially undetermined by context, and are by default determined by the subject. Consider-statements, however, are simply counterstance contingent, because the speaker is expressing a proposition that necessarily relies on context, but fails to be objective. In short, consider-statements are the subjective expression of a fact that relies on context, not on the subject themselves.

As we said, moral predicates can also be used when the speaker does not have a direct experience of a situation, but instead knowledge of some state of the world. In 2., Marc couldn't assert this if he didn't know abortions existed. The difference relies then in the relation the speaker holds with this state of the world. Going back to 1. and 2., let us look precisely at how they differ. Krifka (2012) gives an account of two different ways in which a speaker can make a general statement, which provide some insight for the relation

the speaker holds with the world. In what follows, we don't aim at giving an account of the way general statements behave in relation to moral predicates (i.a Cohen 2010), but we will borrow Krifka's distinction to explain how the relation the speaker holds to the world can account for moral predicates.

A speaker that assumes language is fixed can communicate about the world. In 1., the speaker assumes that her addressee knows what something being tasty means, and according to this describes her experience with sea urchin as tasty. This is a case of what Krifka would call a "description". The speaker, when using moral predicates, is communicating about the language she uses to define the world. In saying 2., Marc is stating that according to him, abortion falls into the category of things that are wrong. This is what Krifka calls a "definition". Marc is defining abortions as wrong, and even if someone might disagree with him, at this stage of the conversation he is redefining wrongness by adding abortions into the set of wrong things.

This definition is stronger than a mere description of facts, and act as if Marc was declaring a new true fact about the world. In the literature, moral predicates were often described using expressivism (i.a. Hay 2011, Marques 2016, Ruiz & Stojanovic 2019) the meta-ethical view according to which moral claims aren't truth-conditional, and simply express the non-cognitive state in which an individual is in. This view is useful to describe taste for example, since someone finding sea urchin tasty is simply describing her experience, not aiming at truth. Moral predicates however are used by a speaker who is acting as if her statement about the world should become the new norm of thought.

A form of expressivism developed by Blackburn (1993) is making the hypothesis according to which, when expressing moral claims, a speaker is not just expressing an attitude, but is also acting as if this claim was true in the real world. Moral claims aren't true or false ; we cannot universally categorize abortions as wrong. However, in the case of moral claims, the speaker is defining a fact from the world as wrong, and acting as if this definition is true. This philosophical claim about the 'quasi-truth' stance a speaker is taking when expressing moral claims could be helpful to give a finer grained understanding or the differences between expression of taste (purely subjective) and factual description of the world (purely objective). Moral claims indeed, function as definitions, holding the place of an intermediate category, between pure description of the world and pure expression of attitudes.

Fink, Sascha Benjamin (Otto-von-Guericke-University Magdeburg)

Activation Bias

Neural activation is currently given the limelight in nearly all aspects of neuroscience and neuropsychology. For example, most theories of consciousness focus on activation as that which brings about consciousness. Activations in specific areas, as predicted by some theory X, are considered evidence for X (Yaron et al. 2022), e.g. task-related activation in the prefrontal cortex is seen as evidence for the Higher-Order-Though-Theory (Lau & Rosenthal 2011). Or an increase in activation in specific brain areas is seen as an indicator of the locus of a neural correlate of consciousness. Similarly for other types of psychological processes. Consider also the common "reverse inference fallacy": In arguing whether some psychological process X is involved in a different psychological process Y, it is mainly activation – not inactivation – in an area associated with X is seen as evidence for X contributing to Y. Activation is where the action is.

Is this concentration on activation justified? Here, I question this narrow focus on activation. I argue that it is possibly a form of malign "activation bias". Why? Because it is highly likely that other types of neural events like inactivation, decreased activation, inhibition, and task-independent default mode activation – i.e. task-unassociated neural default or reduced activations (TUNDRA) – play an equally important role for consciousness. TUNDRAs deserve a share of the limelight.

Several arguments can be made to emphasize the importance of TUNDRAs, based on very different premises, e.g. based on a common understanding of a "neural correlate", on methodological considerations, or on widespread theories in the philosophy of mind. For example: (1) Correlation is, statistically, simply the dependence of two variables. While we prefer a positive correlation with activation, a negative correlation is equally possible. It could then be that the instantiation of a psychological process correlates highest

with a mix between activation in some and inactivation in other areas. (2) In functionalism, a psychological process is partly individuated by its effects. But in order for area A to affect more remote areas (e.g. visual areas having specific effects on prefrontal areas), TUNDRAs in other areas are crucial because then the TUNDRA-areas interfere less with the causal potential of A-activation. TUNDRAs then co-determine whether activation in A is indicative of some psychological process, is involved in bringing about consciousness, and so on. Focusing on A-activation alone fails to acknowledge the importance of non-interference with A's causal potential. (3) In information-based theories of, e.g., consciousness (like IIT) psychological states are associated with patterns of informational values which are, in turn, associated with activation values. As Tononi (2004) stressed, inactivations also carry information and therefore may contribute to consciousness. Informational theories of the mind then naturally give TUNDRA a function to play. (4) Holistic accounts like neurophenomenal structuralism (Fink et al. 2021, Lyre 2022) or connectionism (and associated methods like RSA, see Kriegskorte et al. 2008) identify conscious experience with structures in neural activation spaces. Here, inactivation contributes to the specifics of these structures.

In conclusion, there is no principled reason to discard TUNDRAs as being crucial for psychological states or consciousness. A stronger focus on TUNDRAs is then prudent. This line of argument differs from other critiques of focusing on activation, e.g., Block (in Peters et al 2017) suggests that activation does not differentiate conscious from unconscious perception. There, the focus is still on activation, but its unique relation to conscious processing specifically is doubted. Here, the focus on activation itself is put into question. The argument can be read as contrary to Block's point: Inaction need not be uniquely associated with unconsciousness.

If activation bias is unjustified, then this affects methodology. For example, subtracting resting state data from task-related activation in order to find NCCs becomes problematic. Additionally, studies on default mode networks deserve more prominence. I end with such suggestions for future research.

Sources

Fink, S. B., Kob, L., & Lyre, H. (2021). A structural constraint on neural correlates of consciousness. *Philosophy and the Mind Sciences*, 2.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 4.

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in cognitive sciences*, 15(8), 365-373.

Lyre, H. (2022). Neurophenomenal structuralism. A philosophical agenda for a structuralist neuroscience of consciousness. *Neuroscience of Consciousness*, 2022(1), niac012.

Peters, M. A., Kentridge, R. W., Phillips, I., & Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of consciousness*, 3(1).

Tononi, G. (2004). An information integration theory of consciousness. *BMC neuroscience*, 5, 1-22.

Yaron, I., Melloni, L., Pitts, M., & Mudrik, L. (2022). The ConTraSt database for analysing and comparing empirical studies of consciousness theories. *Nature Human Behaviour*, 6(4), 593-604.

Fitzgibbon, Lily (University Of Stirling), Zoe Ryan (University of Reading) and Helen Dodd (University of Exeter)

Individual differences in preferences for uncertainty

Introduction

Individual differences in responses to uncertainty are increasingly recognised as important predictors of both learning and wellbeing. For example, curiosity has been shown to rival the influence of intelligence on academic performance (von Stumm et al., 2011) and is associated with life-satisfaction and engagement in growth-oriented activities (Kashdan & Steger, 2007). In contrast, Intolerance of Uncertainty (IU) is a transdiagnostic marker of anxiety in both children and adults, and by definition suggests negative affective responses to uncertain situations. It is often theorised that curiosity entails positive and approach-oriented responses to uncertain situations, that in turn lead to learning opportunities and mastery of challenging

materials (Gallagher & Lopez, 2007; Jirout & Klahr, 2012). However, little is known about how differences in these trait measures manifest in children and adults' feelings and behaviours when faced with uncertain situations, and whether these traits confer advantages or disadvantages when it comes to learning under uncertainty.

In two independent studies, we investigated how trait measures of IU and curiosity related to children's (Study 1) and adults' (Study 2) behaviour and affective responses under varying levels of uncertainty. In Study 2, we additionally examined whether trait intolerance of uncertainty and curiosity moderated the memory advantage that is conferred by prolonged uncertainty (Gruber et al., 2014).

Study 1

In the first study, we examined whether individual differences in responses to uncertainty predict children's information seeking and affective responses under different levels of uncertainty.

Method

133 children aged between 8 and 12 years old completed a computerised button-pressing task adapted from the adult curiosity literature (Hsee & Ruan, 2016). On each trial, children were presented with an array of buttons that produced sounds when pressed. In the low uncertainty condition, the majority of the buttons were labelled to reveal whether the buttons would produce neutral or aversive sounds. In the high uncertainty condition, the majority of the buttons were labelled with a '?', so it was unknown whether they would produce a neutral or aversive sound. Prior to being allowed to press the buttons, children's emotional responses to the arrays were recorded in two ways: self-report and webcam recordings of their faces. Parents completed trait measures of their children's curiosity and IU.

Results

We found that children pressed more buttons in the high uncertainty condition than the low uncertainty condition, replicating Hsee and Ruan's (2016) findings with adults. We also found that children rated higher in trait curiosity were happier overall during the task, although this was not moderated by uncertainty. No other findings were significant, suggesting that trait measures of curiosity and intolerance of uncertainty were not related to children's information seeking or affective responses under uncertainty.

Study 2

In the second study, we examined whether individual differences in responses to uncertainty predict adults' preferences for remaining in a prolonged state of visual uncertainty. Further, we examined whether the advantage that uncertainty can confer for memory was moderated by individual differences in responses to uncertainty.

Method

Seventy adult participants completed a task in which blurred pictures were revealed either immediately or slowly, adapted from Jepma et al., (2012). In two counterbalanced blocks, participants pressed a key on the keyboard to reveal the images immediately or slowly. In the slow reveal block, participants experienced prolonged uncertainty as the image revealed gradually over 10 seconds. In the immediate reveal block, the image was revealed immediately upon the button press, and then gradually returned to being blurred, thus controlling the visual input between the trials. As a measure of affective response to the uncertainty manipulation, participants rated how they felt after each block. In the third block, participants could choose how the images were revealed. Participants also completed trait measures of curiosity and IU. Memory for the images was assessed by a free recall test in which participants listed all the images they saw.

Results

Neither trait curiosity nor IU predicted choices for prolonged states of uncertainty (slow image reveal) or participants' affective responses to uncertainty. Replicating previous findings, participants remembered more images from the prolonged uncertainty block. However, this was unrelated to individual differences in curiosity, IU, and preferences for prolonged uncertainty in the choice block.

General discussion

In two studies using different tasks and populations, we found no relationship between trait measures

of curiosity and IU and behavioural and affective responses to uncertainty. This raises several important questions about the relationship between trait measures and behaviour, and is in line with other recent studies demonstrating little connection between self-reported measures of personality traits and performance on tasks purporting to address the same constructs (e.g., Eisenberg et al., 2019).

Gadsby, Stephen (Antwerp University)

Tattoos as social signals

We often use tattoos to infer people's personalities or moral character. For example, those with tattoos are rated as less trustworthy and intelligent (Degelman & Price, 2002; Timming & Perrett, 2016) and are less likely to receive offers when posting second-hand items for sale online (Doleac & Stein, 2013). One way of explaining this phenomenon is with the claim that tattoos are social signals: tattoo owners choose them to communicate evidence about themselves, and others simply respond to this evidence.

There are, however, puzzles related to the idea that tattoos are social signals. For example, while people report using tattoos to infer character traits, tattoo owners often deny the relevance of signalling to their tattoo choices, instead, claiming that their tattoos are motivated by aesthetic preferences. One way to explain this is that tattoos are cues rather than signals—they transmit information but aren't selected for that purpose. A contrasting explanation is that tattoo owners are unaware of the true motivations behind their tattoo choices.

This talk explores the idea of tattoos as social signals and draws out implications for the psychology and philosophy of tattoos. First, I address the question of whether tattoos are cues or signals, through an analysis of the costliness of tattoos. Good signals are costly. Telling everyone that you are wealthy is a bad way to signal your wealth (it may even have the opposite effect). Anyone can claim that they are wealthy, it costs nothing. Wearing expensive clothes or driving expensive cars, on the other hand, is a good signal of wealth—in most cases, only genuinely wealthy people can bear the costs of such behaviour.

Those who note the social signalling role of tattoos generally refer to costs like time, money, pain, and health risks (Wohlrab, et al., 2017). However, the costliness of a signal should relate to the quality being signalled (Zahazi & Zahavi, 1999). Expensive clothes are a good signal of wealth because the relevant cost is financial. This principle, I argue, helps to constrain the qualities that tattoos can be used to signal and illuminates a signalling role that tattoos are particularly well-suited to, namely, commitment to a social group. I show how consideration of the costliness of signals helps to distinguish between tattoos that are cues and those that may be social signals, driven by unconscious motives. I argue that when tattoos are costly in the right way, there is good reason to assume that owners who deny their signalling role are confabulating.

Building on the idea of tattoos as social signals, I also address the moral dimension of tattoos, introducing the idea that tattoos can act as self-signals. The standard argument for self-signalling begins with the observation that many of our character traits are hidden from us. Consequently, we must infer them based on our observable characteristics (Holton, 2016). For example, when participants are told that being able to hold one's arm in cold water for longer is a sign of a strong heart, they hold their arm in the water for longer (Quattrone & Tversky, 1985). This is explained with reference to the idea that participants signal to themselves that they possess desirable (but hidden) traits, in this case, physical health. Just as participants hold their arms in cold water to signal healthiness to themselves, one's tattoos can self-signal other desirable attributes. A skull and crossbones tattoo, for example, could self-signal toughness.

Tattoos represent a unique form of self-signalling, in virtue of being pre-committing: tattoos constrain our future selves in a way that putting one's hand in cold water doesn't. Specifically, they guide and constrain the values we adapt and the people we become. In getting a swastika tattoo, for example, one increases the chances of adopting or maintaining the values associated with that symbol. This illustrates an important philosophical aspect of tattoos, namely, that they can be used as morally aspirational precommitment signals (MAPS), which guide us towards becoming (or remaining) the kinds of people that we aspire to be. I develop this idea and outline some of its moral implications.

This talk illuminates several puzzles related to the philosophy and psychology of tattoos. Specifically, it makes progress on distinguishing tattoos that are social signals (driven by unconscious motives) and those that are cues (transmitting unintended information). It also introduces an overlooked moral value of tattoos, namely, their ability to function as MAPS.

Garbisch, Isa (University of Göttingen), Joana Lonquich (University of Göttingen), Marina Proft (University of Göttingen) and Hannes Rakoczy (University of Göttingen)

Children's understanding of subjective intentions in a retrospective event interpretation task

Background

How do children develop an understanding of the subjective reasons and intentions with which rational agents act? Intentions can be subjective in the sense that an agent does many things simultaneously ("travelling to an island", "travelling to the treasures") but the intentionality depends on the aspect or description under which she acts (e.g., Anscombe, 1957; Goldman, 1970; Searle, 1983). For example, if a pirate travels to an island where he mistakenly believes a coin to be hidden, he is intentionally traveling to that island, but not intentionally traveling to an empty treasure chest. To make sense of his action, one must represent that his false belief about the location of the coin led him to his intentional action. Children around 4-5 years can master action prediction (e.g., "where will the pirate go now?") or belief ascriptions (e.g. "where does the pirate think the coin is?") tasks (such as standard explicit false belief tasks). But when children are directly asked whether the pirate went to the empty treasure chest on purpose, they have problems with these kinds of subjective intentionality questions until they are about 5-8 years old (Kamawar & Olson, 2011; Proft et al., 2019; Schünemann et al., 2021). It remains unclear what the source of these difficulties for children may be. One interpretation is that children misunderstand the (de re/de dicto) reading of the question (Schünemann et al., 2021): Understood extensionally, the pirate did travel on purpose, and he did end up what turned out the be the empty treasure chest; but on an intensional reading, the pirate did do something on purpose (travel), but not do this particular action (travel to the empty chest). Now, children may be able to understand this conceptually in principle, yet, for linguistic or other reasons, may not spontaneously find de dicto readings obvious.

Rationale and Method

Therefore, the rationale of the present study was to reduce such potential linguistic performance factors that may mask children's conceptual competence. Rather than asking directly whether a given action was done on purpose under a given description, we used an indirect way to measure children's subjective intention understanding. Children first heard about an agent and his subjective beliefs and desires. For example, the pirate wants an object X and believes it is on island A. Then they heard about an outcome of the story that was either consistent with the desires and subjective beliefs of the pirate (pirate is now at island A) or inconsistent (pirate is now at island B). The story was set up in such a way that it was ambiguous what happened in between: did the pirate actively travel there, or was his boat passively transported there by the wind in the course of a storm? The crucial test question was thus whether the pirate ended up where he was by actively going there himself or by being blown by the wind.

In test conditions the pirate could be wrong about the treasures in two ways. Either he held a false belief about the location of an object (change-of-location: the treasure was moved from one island to another in his absence) or mis-represented the identity of an object (dual-identity: he knew there was an object on a given island, but did not know it was a treasure). In control conditions, the pirate was fully aware of the objects' locations and identities.

In a mixed design (between: age; within: condition, event type), children aged 4 to 7 years each saw 4 training trials (without belief manipulations) and 6 test trials: 2 per condition (change-of-location, control, dual-identity), 1 consistent and 1 inconsistent event type. After each trial, children judged whether the outcome was brought about by the agent or by external factors.

Preliminary results

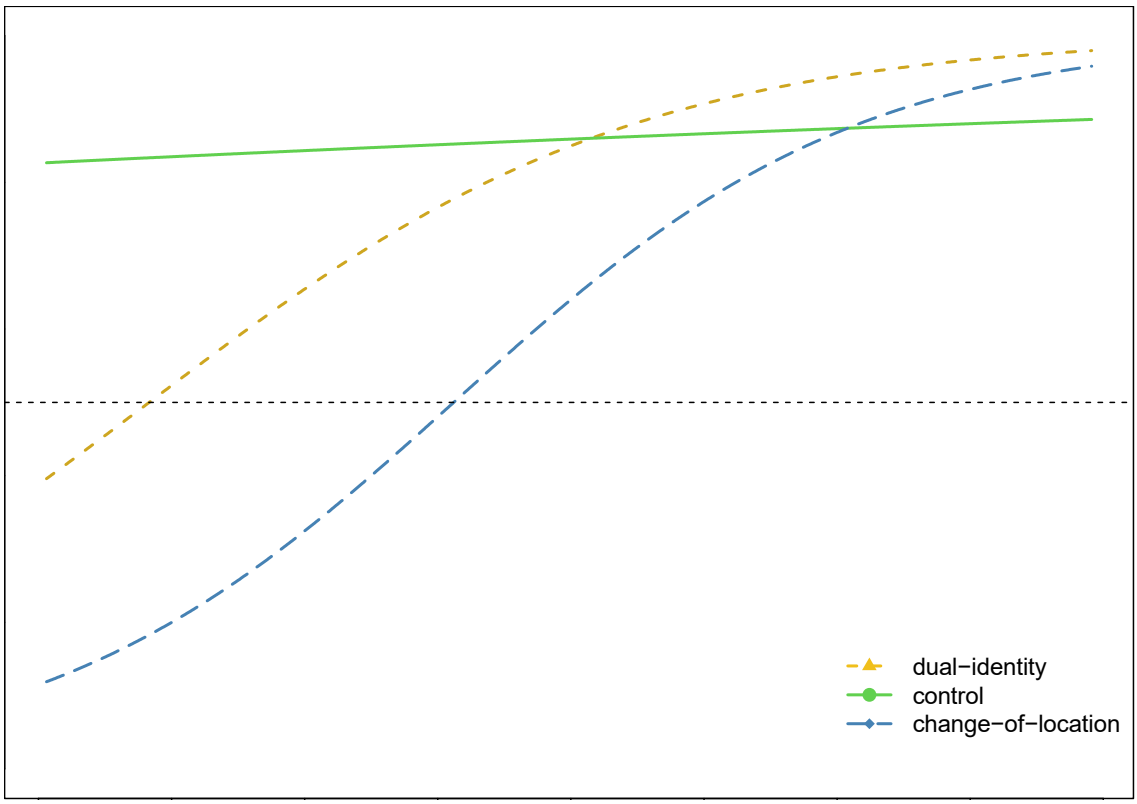
The preliminary results of the preregistered study (https://doi.org/10.17605/OSF.IO/79N8T) with N = 99 (of N = 120) children are the following: Children before the age of 5-6 years made the wrong ascriptions in test conditions (change-of-location and dual-identity). Younger children, but not older children, were more likely to say that the agent either intentionally drove to the objectively better island (the one with the treasure about which, however, the agent did not know) or was driven by the wind to the objectively worse island (the one regarding which the pirate mistakenly believed that there was a treasure). In contrast, in control conditions, in which the agent had no false belief about the treasures locations or identities, children across all ages made correct intentionality judgments: whether the pirate drove to the island on his own or not was purely based on the consistency of the outcome and his other mental states. Interestingly, children across all ages performed significantly worse in the test conditions in which the agent had a false belief about the object's location than when he mis-represented its identity. Likewise, younger children were significantly below chance level in the first test conditions and at chance in the latter.

Conclusion

The present findings converge with previous studies in suggesting that younger children seem to have systematic difficulty keeping apart agents' desires and intentions (see Astington, 2001; Schult, 2002). These difficulties might not be related to linguistic performance factors of the question. The present findings suggest that understanding subjective intentions develops only later and might be a case of an advanced theory of mind ability.

Figure 1

Probability of correct responses as a function of age, condition and their interaction. Correct responses mean that children answered in consistent trials (outcome matches the pirates subjective beliefs and desires) that the pirate drove there on his own and that the wind drove him there in inconsistent trials. Dots represent individual responses per condition. Lines are fitted values based on a binomial mixed effects model. Colored areas around indicated 95% confidence intervals. Fitted values and their confidence intervals have been obtained via bootstrapping with 10,000 boots. Chance level = 50%.



Gauker, Christopher (University of Salzburg)

Amodal Completion: Imagination or 3D Modeling?

When we visually perceive a scene, we may in some sense represent occluded parts of the scene that do not reflect light into our eyes. In psychological research this is called amodal completion. Amodal completion may be based on principles of good form, shape symmetries or even acquired knowledge about the kinds of things perceived. But in what sense do we represent the occluded elements? Bence Nanay has argued that amodal completion takes the form of representations in mental imagery of the occluded portions (2010, 2018). I will argue that Nanay's account requires us to imagine the unimaginable. It requires us to imagine the occluded portions in precisely the places in the scene where they are hidden from view. An alternative account begins with the observation that on the basis of our perception we may build a three-dimensional mental model of the spatial structure of the scene perceived. It is such 3D models that allow us to imagine what an object would look like if we viewed it from the other side. Amodal completion consists in building such a 3D model, which we can do without imagining what the scene would look like from a different perspective.

Gouveia, Steven S. (Mind, Language and Action Group, Institute of Philosophy, University of Porto)

Philosophy and Neuroscience: An Experimental Philosophy Study

One of the central issues in Philosophy of Neuroscience and Neurophilosophy is concerned with the (meta)philosophical question of considering if Philosophy and Neuroscience should cooperate or not together to solve several philosophical problems related with the mind and the brain, and, if so, how this cooperation should take place.

Typically, there are four approaches that relate (or not) the philosophical inquiry with the neuroscientific research: (1) the Isolationist Approach denies any kind of interaction between the two disciplines. Next, rejecting this isolated flavour and arguing for a relational approach, (2) the Reductionist Approach (RA) focuses on specifying how philosophical work can be reduced to neuroscientific work. Arguing against the reductive aspect of (RA), but maintaining a "relational" relationship, (3) the Neurophenomenological Approach (NA) demonstrates the importance of an embodied approach to the study of the conscious mind. Finally, (4) the Non-Reductive Neurophilosophical Approach (NRNA) exposes a particular methodology that conceives the epistemic utility of both philosophical and neuroscientific work.

The goal of our talk is to present the data collected by an xPhil Survey focused on this specific issue. The sample is composed of 230 professional philosophers and neuroscientists (65.7% with a philosophical training and the rest with a neuroscientific/psychology/cognitive science backgrounds). Here is a summary of the different results:

--> Most researchers consider that philosophy should directly or indirectly collaborate with Neuroscience while a minority considers that Philosophy should not engage with Neuroscience;

--> When this (collaboration) is considered, most researchers assume that Philosophy and Neuroscience should collaborate in equal measure, by being conceptually and empirically accurate and plausible, as opposed to the thesis that Philosophy should guide holistically Neuroscience and that Philosophy should be reduced to Neuroscience;

--> The majority of researchers favor naturalism (as opposed to a minority who advocates non-naturalism);

--> More than half of researchers think that an Anti-Cognitivist Approach regarding the definition of Philosophy (the aim of Philosophy is what does or does not make sense, is understanding) is true, against a minority that thinks that a Cognitivist Approach (aim of Philosophy is what is true or false) is true.

This paper will present for the first time in the literature data regarding the different philosophical, methodological and metaphilosophical assumptions used by professional researchers that work at the intersection between Philosophy and Neuroscience. Our goal will also be to find out if there is any strong significance between the different questions analyzed: for example, if the metaphilosophical assumption of the research influence her/his position regarding the relationship between Philosophy and Neuroscience. We will provide some interesting correlations and respective interpretations for the issue at stake.

Grad, Paweł (Faculty of Philosophy, University of Warsaw)

Realism and the Rudimentary Theory of Perception

Consider two theses about the perceptual experience:

Objectivity: a subject's S perceptual experience presents her with mind-independent object x.

Source: S has thought of x as mind-independent thought in virtue of having perceptual experience of x.

My concern in this paper is the controversy between the defenders of Objectivity and Source about the relation between these claims (Evans 1985; Campbell & Cassam 2014; Eilan 2017, Brewer 2020). The main question in the debate is how neatly the objectivity of perception, i.e. the fact that perception presents S with mind-independent physical objects, supports S's representation of x as mind-independent, and more specifically: her realist belief that x is mind-independent.

In this paper, I argue that Objectivity alone cannot explain the role of perception that it plays according to Source. I defend an intellectualist view that the subject's perceptual experience may be a basis of her realist belief in the mind-independent existence of perceived objects only if the subject has the Rudimentary Theory of Perception (RTP) that makes the transition from perceptual experience to the realist belief valid.

My argumentative strategy involves (1) distinguishing between thin and thick objectivity, (2) unpacking the idea of the source of realist belief, (3) presenting an argument for RTP from the thinness of perceptual experience, and (4) identifying objections to my proposal.

1. Strawson's (1979) and Evans' (1985) intellectualist position entails not only a claim about subjective prerequisites of grasping a realist conception of the world (Source) but also a claim that perceptual experience might be objective only if the subject possesses and employs in perception the concept of a mind-independent object. This view ascribes thick objectivity to perception.

ObjectivityTHICK: S's perception of x presents mind-independent x to S as mind-independent.

Tyler Burge (2010) argues against Strawson's and Evans' intellectualism by claiming that causal dependence of perceptions on the mind-independent environment secures perceptual objectivity even if a subject does not represent objects of perception as mind-independent. Burge does not agree with Strawson and Evans not only on what the prerequisites of objective perceptual presentation are but also on how thick is the objective import of perception. More specifically, Burge (2009) claims that:

ObjectivityTHIN: S's perception of x presents mind-independent x to S, but not as mind-independent.

2. Source expresses access internalist intuitions about the relation between the subject's perceptual experience and her realist belief. The idea is that if the relation is to be non-accidental from the subjective perspective, then the transition from the perceptual experience of x to the mental representation (e.g., thought, belief) of x as mind-independent must be normatively robust in the following way: what is presented in perception validates subject's thought that x is mind-independent. As such, Source does not concern concept acquisition by S but rather S's employment of a concept MIND-INDEPENDENT (Cassam 2011).

3. My argument for RTP starts from endorsing ObjectivityTHIN. But thin objectivity of perception is too thin to make the immediate transition from experience to realist belief valid (the content of experience does not support the content of belief). The validity can be secured only by mental representations additional to perceptual experience.

The argument from thinness of experience

ObjectivityTHIN	S's perceptual experience of x (ε) presents mind-independent x to S, but not as mind-independent.
Source	ε is the source of S's representation (R) of x as mind-independent.
Validity	If ObjectivityTHIN and Source, then among S's mental contents are contents < C1,..., Cn > such that they make S's transition from ε to R valid.
RTP:	Among S's mental contents are propositions < C1,..., Cn > such that they make S's transition from ε to R valid.

< C1,..., Cn > are contents of a "rudimentary theory of perception". The very term comes from Gareth Evans (1985, pp 261–2). Understanding RTP as a theory in the full-fledged sense would be a little excessive. The idea is rather that a realist belief is a result of a transition from thinly objective perceptual experience supplemented with a set of additional representations which link logically what is given in perception with the proposition that x exists as a mind-independent object. These additional representations concern general conditions for perceiving an object in a way which entails that the object is mind-independent. This set of representations can be called the "theory of perception" because they explain certain aspects of perception (e.g., the spatial aspect of perceptual representation, or the ability to re-identify the object). The theory is "rudimentary" since it can be mastered by all perceivers equipped with a suitable conceptual apparatus, i.e. by virtually all mature human beings. Another point of contrast with scientific theory is that it can be used tacitly, and not only in the course of a deliberate, explicit theoretical inquiry.

4. The argument is valid but its soundness is open to objections. There are both representationalist (Strawson 1979; Siegel 2010, ch. 7; Spener 2012; Textor 2019) and relational (Campbell & Cassam 2014 ch. 5; Eilan 2017; Brewer 2020) arguments against ObjectivityTHIN. My response, in general, is that neither cognitive science (Feldman 2003; Burge 2009; Peters 2020) nor phenomenology (Mackie 2020) support ObjectivityTHICK. I put aside the doubts concerning the general form of Objectivity and Source since my primary concern is not whether Objectivity and Source are true, but rather how we should understand the nature of the transition from the objective import of perception according to Source. Validity might be criticized both for undermining the character of perception as the unique source of realist belief (Brewer 2020) and for over-intellectualizing the mental processes in question (Burge 2010; Brewer 2020). In response, I reject the uniqueness condition for Source and argue that Validity follows directly from the nature of mental transition implied by Source (Fodor 1975; Dunn & Mandelbaum 2017).

References

Brewer, B. (2020). Perception of continued existence unperceived. *Philosophical Issues* 30, 24–38.

Burge, T. (2009). Perceptual objectivity. *The Philosophical Review* 118, 285–324.

Burge, T. (2010). *Origins of objectivity*. Oxford: Oxford University Press.

Campbell, J. & Cassam, Q. (2014). *Berkeley's Puzzle: What Does Experience Teach Us?* Oxford: Oxford University Press.

Cassam, Q. (2011). Tackling Berkeley's Puzzle. In: Roessler, J., Lerman, H., & Eilan, N., *Perception, Causation, and Objectivity* (pp. 18–34). Oxford: Oxford University Press.

Eilan, N. (2017). Perceptual objectivity and consciousness: A relational response to Burge's challenge. *Topoi 36*, 287–298.

Evans, G. (1985). Things Without the Mind—A Commentary upon Chapter Two of Strawson's *Individuals*. In: Evans, G. *Collected Papers* (pp. 249–90). Oxford: Clarendon Press.

Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences* 7, 252–256.

Fodor, J. (1975). *The Language of Thought*, Cambridge, MA: The MIT Press.

Mackie, P. (2020). Perception, Mind-Independence, and Berkeley. *Australasian Journal of Philosophy* 98, 449–464.

Peters, U. (2020). Objectivity, perceptual constancy, and teleology in young children. *Mind and Language* 37, 975–992.

Quilty-Dunn & Mandelbaum, E. (2018). Inferential transitions. *Australasian Journal of Philosophy* 96, 532–547.

Siegel, S. (2010). *The Contents of Visual Experience*. Oxford: Oxford University Press.

Spener, M. (2012). Mind-Independence and Visual Phenomenology. In: Smithies, D. & Stoljar, D. (eds.), *Introspection and Consciousness* (pp. 381–403). Oxford University Press.

Strawson, P.F. (1979). *Perception and its Objects*. In: Macdonald, G.F. (eds.), *Perception and Identity* (pp. 41–60). London: Palgrave.

Textor, M. (2019). Objectivity and the limits of perception. Phenomenology and the Cognitive Sciences, 18, 879–892.

Grünbaum, Thor (Philosophy, University of Copenhagen)

Temporally extended agency, intentions, and memory

Let us start by getting the phenomenon into view. In the morning, I decide to stop by the Chinese supermarket on my way home from work in the evening. I need noodles for tonight's dinner. Then my workday starts, I give the plan only little thought. But at 11:50 my brother calls me to ask if I can stop by the HiFi-shop on my way home. I remember my plan to stop by the supermarket and say no to the new plan. Opening times allow me to do only one of them. By the end of the day, as I prepare to leave for home, I recall my intention to stop by the Chinese supermarket. I now plan the specific route that will take me to the shop on my way home. Finally, I depart.

Cases like this require that the decision-maker's intention can somehow stay with the agent over extended periods of time. The agent decides at t1 to φ at t4, at t2 they recall their intention to φ while considering other plans, at t3 they recall their intention to φ while reasoning about how to φ, and at t4 they engage in performing φ. Importantly, the agent has to remember their intention to φ not only when the time to perform the action has come, but also when considering potentially conflicting plans and coordinating with others (t2) and when engaging in reasoning about means to the end (t3).

In this talk, I argue that this form of temporally extended agency is supported by a special form of memory. When the agent remembers their intention to φ, it is important that they can recall their intention without starting to reconsider whether to φ. When becoming aware of their intention because they are considering other plans (t2) or engage in means-end planning (t3), it is important that they can become aware of their intention in such way that their remembering does not invite considerations about whether to do the action at all. This type of default non-reconsideration constraint on remembering one's intention cannot be satisfied by standard forms of episodic memory, or so I argue in this talk. If our agent is to satisfy basic constraints on temporally extended agency, as understood by a planning theory of intention, we must make room in our mental ontology for a special form of memory for intention.

In this talk, I argue that human temporally extended agency is made possible by a form of memory which is characterized by the following two features:

a) Intentions can persist in long-term memory as real standing intentions.

b) Remembering one's intention can have the form of a simple re-activation of the intention.

Each of these requirements is a substantial claim. They involve contentious empirical and theoretical implications that are in conflict with popular views on intentions and temporally extended agency.

(a) is denied by any theory claiming that, in the normal case, an agent has a long-term persisting intention to φ if (i) they formed the intention to φ, (ii) they can remember and thereby occurrently believe that they formed the intention to φ, and (iii) the remembering would cause them to form the intention to φ again. That is, (a) is denied by accounts that claim that in the normal case intentions have long-term persistence by being encoded into memory in the form of beliefs about the original intention formation. This is form of anti-realism about standing intentions. To say that a person non-occurrently intends to φ – for instance, to say about a sleeping person that they intend to go shopping when they wake up – is to say that the intention is stored in their memory as a belief about their decision to go shopping and that this belief provides them with a disposition to decide to go shopping.

(b) is denied by any theory claiming that (i) memory for intentions is a form of episodic memory, and (ii) a person has episodic memories about themselves if and only if they can form beliefs about themselves based on having been in the experiential situation in the past. The problem is this. If a person is remembering their intention to φ if and only if they are occurrently believing that they formed the intention to φ (where this belief is based on awareness involved in forming the intention), then we should expect that agents often remember now their intention φ without intending now to φ. After all, on this account, remembering is to have a belief about a past self. By contrast, if the normal case of remembering one's intention is a simple re-activation of the intention, then remembering now one's intention to φ guarantees intending now to φ.

I argue for each of the requirements (a) and (b). Their rejection has important implications for our explanation of the target phenomenon, temporally extended agency. I will proceed as follows. First, I articulate basic assumptions concerning persisting intentions and memory and use the assumptions to pitch two models of memory for intentions against each other. According to the first model (the anti-

realist model), standing intentions do not really exist and remembering one's intention is a form of episodic memory retrieval. According to the second model (the re-activation model), standing intentions really exist and remembering one's intention is the process by which a standing intention is re-activated. Then I proceed to describe a number of features of temporally extended agency that an account of the role of memory in temporally extended agency should be able to explain. They all relate to a basic non-reconsideration constraint on planning agency. Finally, I argue that the re-activation model but not the anti-realist model is able to provide a convincing account of the three features. Only the re-activation model can satisfactorily explain how non-reconsideration can be a default feature of temporally extended agency.

Heemskerk, Johannes (University of Warwick)

Theory or Gloss: Can We Extract a Theory of Content from Cognitive Science?

Naturalistic theories of representational content remain one of the central concerns of Philosophy of Mind. In this paper I consider a common methodology used by contemporary theorists, and consider an objection to this methodology, based on an argument due to Frances Egan. The methodology involves analysing cognitive science, with the aim of extracting an implicit theory of content determination. As Tyler Burge writes, the methodology is based on the thought that cognitive science has discovered, "without being fully aware of its own accomplishment" [Burge, 2010], a naturalistic theory of content. Egan's challenge consists in her argument that content talk is a 'gloss' designed to aid comprehension of mathematical functions carried out by the brain.

As such, content attributions in scientific theories are made on the basis of pragmatic decisions such as how to best communicate the theory to readers. Since there is no implicit naturalistic theory of content in cognitive science, none can be extracted.

Given the ubiquity of the above methodology, Egan's argument threatens to undermine much of the contemporary content naturalisation project. In this paper I provide an analysis of influential work in cognitive science, focusing in particular on a study on facial recognition [Chang and Tsao, 2017]. I argue, contra Egan, that the representational content invoked in this and other studies meets Egan's two criteria for inclusion in the theory proper. These criteria are that content should be 'essential' and determined by a 'privileged naturalistic relation' [Egan, 2018]. I describe both criteria below. First, content is 'essential' to the representational states and structures under consideration. By essential, Egan means that representations are individuated by their content. Through an analysis of the cognitive science literature I attempt to show that the authors do in fact individuate representations by way of their content. This is no accident, but central to the explanations on offer, since the relevant studies' explananda involve the cognitive system's interaction with its environment.

Second, content is determined by a privileged naturalistic relation holding between a representation and its distal content. In the case study I focus on, this naturalistic relation is one of 'encoding'. Encoding is not itself a mere gloss, but comes with hefty theoretical commitments which are central to the explanation provided. This generalises, and it emerges that cognitive scientists are implicitly invoking a theory of content determination which draws heavily on the resources of communication theory.

I argue that we must concede to Egan that unless content is used in the theory proper in her sense, it is not informed by an implicit naturalistic theory. To this end, I close by suggesting a general schema by which to identify whether content is used within the theory proper of cognitive science; the content must be specified by way of a technical concept, the invoked relationship between the representation and a distal item must be theoretically rich, there must be a possible reduction to a naturalistically acceptable language, and the content must be explanatory of some capacity of the system in a way which 'makes a difference' to the system itself. It is the work of philosophers to construct a coherent theory of content drawing from various studies which conform to this schema.

References

[Burge, 2010] Burge, T. (2010). Origins of objectivity. Oxford University Press.

[Chang and Tsao, 2017] Chang, L. and Tsao, D. Y. (2017). The code for facial identity in the primate brain. Cell, 169(6):1013–1028.

[Egan, 2018] Egan, F. (2018). The nature and function of content in computational models. In *The Routledge handbook of the computational mind*, pages 247–258. Routledge.

Hendrickx, Malte (University of Michigan Ann Arbor)

Mental-Effort First: Explaining Effort

Mental and bodily effort play a central philosophical role in explaining the nature of achievement, skill, learning, self-control, and free will. Yet research on effort is scattered across discipline and fields. What are efforts, and how do different kinds of efforts relate to each other?

This paper makes three novel contributions to the existing literature. First, I draw on psychological evidence to argue that mental efforts are the (goal-directed) deployment of executive control. Secondly, I draw on psychological and kinesiological evidence to argue that physical efforts are the (goal-directed) deployment of central motor commands. Yet, when proposing two domain-specific accounts of effort, a question arises as to whether there is a domain-general concept of effort. I argue that there is. Specifically, I argue for a novel mental-first theory of effort by showing that mental efforts are both necessary and sufficient for any effort. Consequently, I propose a novel theory that explains general effort as the deployment of executive control.

Literature Review

A problem facing any theory of effort is the abundance of concepts to be explained. Agents can make efforts (henceforth: efforts). Efforts are accompanied by feelings of effort.1 Tasks that require effort are difficult. To make matters worse, all three interrelated concepts come in physical (lifting a pumpkin) and mental (doing math) flavors. An unresolved problem in the literature on effort is explaining how these different concepts fit together: is there a general notion of effort? Efforts, feelings of effort, and difficulty have different target profiles. Efforts are things we do, are controlled and goal-directed, and have success conditions. The feelings of effort, on the other hand, are passive, outside of our control, and not goal-directed. Difficulty is neither a thing we do nor a feeling but a relational property of a task and an agent: the more effort a task requires of an agent, the more difficult it is.

I set the stage for my positive proposals by briefly revisiting and refuting current proposals of general theories of effort: resource-based views, feeling-first views, and force views. Resource-based views take efforts to be the expenditure of a depletable resource. Yet I argue that such views face insurmountable empirical worries: converging evidence suggests no such depletable resource exists. Feeling-first views take efforts to be actions accompanied by a feeling of effort. Contradicting these accounts, I show clinical and non-clinical examples of agents that are able to make efforts absent any feeling of effort. Force views take efforts to be exerting a force against a resistive force. I argue that force views explain only physical efforts: mental efforts often do not strive against a resistive force.

Positive Proposals

I then make two positive contributions: the first proposes a new theory of mental effort, and the second proposes a new theory of physical effort. Both accounts arise out of a review of the empirical literature on the respective feelings of mental and physical effort. Specifically, a host of recent work concerning the feeling of mental effort suggests that it arises from an expected value calculation for the deployment of executive control. I argue that this suggests the following:

Mental Effort Thesis: Mental efforts are the deployment of executive control by an agent.

I then review recent work suggesting that the feeling of physical effort arises as a function of central motor commands. I argue that this suggests that physical efforts are best understood this way:

Physical Effort Thesis: Bodily efforts are the deployment of central motor commands.

I argue that both accounts naturally fit the target profile of effort. I further propose that they deal with problem cases that current theories of effort cannot handle.

Having proposed two domain-specific accounts of effort, I turn toward their relation: is there a general genus of effort? I review empirical and philosophical evidence that shows that no effort is possible without mental effort. Mental efforts are thus necessary to all efforts. Furthermore, the deployment of executive control is necessary for any goal-directed action: since efforts are goal-directed actions, any effort must involve the deployment of executive control.

Since mental efforts are efforts, they are sufficient for effort. This relation is not trivial: I argue that any issuing of central motor command requires proportional executive control.

Hence, mental efforts, on the view I propose, are explanatorily and causally prior to physical efforts. Rather than being a distinct form of effort, physical efforts are a mode of mental effort. I call this somewhat surprising result the "mental-first" view of effort, which can be summarized this way:

Mental-First Effort Thesis: The deployment of executive control is necessary and sufficient to all efforts.

I end by sketching some implications of my view for debates surrounding difficulty, learning, and achievement. I show that the mental-first thesis clarifies these debates and dissolves some stubborn puzzles raised in prior philosophical work.

Henke, Benjamin (Antwerp Centre for Philosophical Psychology)

Explaining the Impact of Visual Illusions on Perception and Action: Comparing the Two- and One-Representation Accounts

Milner and Goodale's (2004, 2006) Perception/Action Model proposes two separate cortical visual processing streams: a ventral stream for object identification and further cognitive processing, and a dorsal stream for online motor control. The model contains several sub-theses, including that the visual streams are independent (or mostly independent) systems for perception and action ('Two-Systems') and that consequently perceptual and action operate over distinct representations ('Two-Reps').

A key piece of evidence supporting the model is the finding that visual illusions affect action tasks less than visual perception. For example, Aglioti et al. (1995) found that a 3D version of the Ebbinghaus illusion — where a center circle surrounded by large circles appears smaller than the same circle surrounded by small circles (see Fig. 1) — minimally influences grip aperture when reaching for the center circle. Similar results have been obtained for a wide variety of visual illusions. If illusions have a smaller impact on guidance representations than on conscious perception, then it follows that the representations are not the same. Thus, illusion studies have been considered strong evidence in favor of Two-Reps, and by extension Two-Systems.

This support has been challenged in recent years, however. Thor Grünbaum (2017, 2018) has challenged the inference from illusion studies, suggesting that, despite appearances, distinct performances on perception and action tasks are consistent with a single, common representation ('One-Rep').

In this talk, I reassess the evidence from illusion studies for the Perception/Action Model. After briefly addressing challenges to the experimental results, I focus on evaluating the support these results provide for Two-Reps. I argue that the results are better explained by Two-Reps than One-Rep, contra Grünbaum's argument.

In section 1, I explain the Perception/Action Model, focusing on the distinction between Two-Systems and Two-Reps. This distinction is crucial, as while most treatments, including Grünbaum's, have regarded the confirmation of Two-Reps as tantamount to confirmation of Two-Streams, I will conclude that illusion studies support Two-Reps without supporting Two-Streams.

In section 2, I present the illusion studies, describe their purported support for the Perception/Action Model, and highlight three crucial findings from the illusion results. First, there is evidence that visual illusions are presented in perceptual experience itself, rather than merely inferred from experience. This is suggested, for example, by the fact that perceptual illusions persist despite our knowing that they misrepresent. Second, it has been shown that a wide range of visual illusions have a smaller impact on

action than perception. Crucially for my argument, these different illusions distort distal properties in various ways, such that there is no unique relationship between the veridical and illusory properties. Third, it has been found that visual illusions that do impact action are typically generated early in visual processing, while those that do not (or do to a lesser degree) are thought to be generated late in visual processing (Kozuch 2022).

In section 3, I present Thor Grünbaum's (2017, 2018) argument that, contrary to appearances, illusion studies fail to support Two-Reps. Grünbaum presents two interpretations of the visual streams. According to Milner and Goodale's interpretation, which assumes Two-Reps, the streams generate distinct representations for their respective downstream task. According to an alternative, which assumes One-Rep, the streams respond to a common representation generated early in the visual system. On this view, the role of the visual streams is not to generate representations but to 'transform' the unified representation into precise commands, such as motor commands. Grünbaum claims that any difference in behavior which is explainable by a difference in representation, according to Two-Reps, can also be explained by a difference in transformational processes, according to One-Rep. Thus, the evidence cannot decide between the two models.

In section 4, I argue, contra Grünbaum, that Two-Reps does a better job than One-Rep in explaining the three findings mentioned in section 2.

I first demonstrate that Two-Rep provides a straightforward explanation of the three findings. The first two findings are easily explained by the fact that perception and action employ distinct representations, and thus have different properties. The third finding is explained by the fact that illusions generated before the anatomical divide are likely to have residual impact on action.

I then explore two versions of One-Rep and examine whether they can explain the three findings. The problem I raise occurs for each illusion whose impact differs between perception and action. One version of the One-Rep claims that, for such an illusion, the unified representation represents only the veridical property. This view can explain the second finding, since action appears to respond to an approximately veridical representation. But it is inconsistent with the first and third findings, since it fails to explain the presence of any illusions in visual experience or in action.

A second version of the One-Rep claims that, for an illusion whose impact differs between perception and action, the unified representation represents only the illusory property. This account can explain the first finding, since the illusory property is directly represented in the unified representation. And, if Grünbaum is right about transformational processes, then it can explain the second and third findings by positing transformational properties that correct for the many errors of the unified representation. However, because the second finding is that there are many sorts of differences between the illusory and veridical properties, we would need to posit many sorts of transformational processes to correct for these various illusory properties. And because the third finding is that some illusions impact action while others do not, we must posit such transformational processes for just the right illusions. But while both maneuvers are possible, they each add substantial complexity to the One-Rep view (and the latter maneuver is particularly ad hoc). Thus, even if this version of the One-Rep can explain the illusion findings, I maintain that Two-Rep provides a simpler, more plausible explanation.

I thus conclude that illusion studies continue to support Two-Rep over One-Rep.

Heppell, Marlize (University of the Free State) and Lindi Nel (University of the Free State)

The experiences of adolescent daughters of mothers diagnosed with borderline personality disorder

This study aimed to describe, interpret and gain deep insight into the lived experiences of adolescent (aged 12 – 18) daughters with mothers diagnosed with Borderline Personality Disorder (BPD). Patients diagnosed with BPD present with long-standing impairment in various domains of life, including identity disturbances, interpersonal challenges and difficulties with impulse and affect regulation. The volatility BPD patients experience in their interpersonal relationships is a core feature of the disorder that disrupts patients' professional and personal lives. In probing the personality structure of borderline personality

disorder, the primary concern is the ego-structure pathology which includes nonspecific manifestations of ego weakness (e.g., a lack of anxiety tolerance) as well as defense strategies such as primitive idealization, splitting, excessive projection and projective identification and omnipotence that can be associated with devaluation. Research indicates that childhood adversity and neurobiological and genetic factors may contribute to the disorder's aetiology. Due to the substantial impairment that is intrinsically part of a BPD diagnosis, the question is how this will affect the mothering abilities of women diagnosed with the disorder. Specifically, mothering adolescent daughters in the fifth developmental phase when they are confronted with identity forming versus identity confusion. In this phase, attachment forming has already been established but a revision of the internal working models usually occurs as adolescents are much better equipped than younger children to critically review their internal working models.

When a mother has a self-valued and secure IWM, she will be able to teach her child independence and trust, but when the maternal IWM is based on low maternal expectations and trust, it will probably be carried over from one generation to the other. It was found that the offspring of mothers diagnosed with BPD presented with a significantly higher number of psychiatric diagnoses and obtained much higher scores on a global impairment rating. It became evident from the participant's responses in the interviews that the mother-daughter relationships were compromised, and a lack of trust towards their mothers existed in most cases. Several participants decided to detach from their mothers as they found the inconsistencies, projection, confusing boundaries, and lack of emotional support to make them feel unsafe in their mother's company.

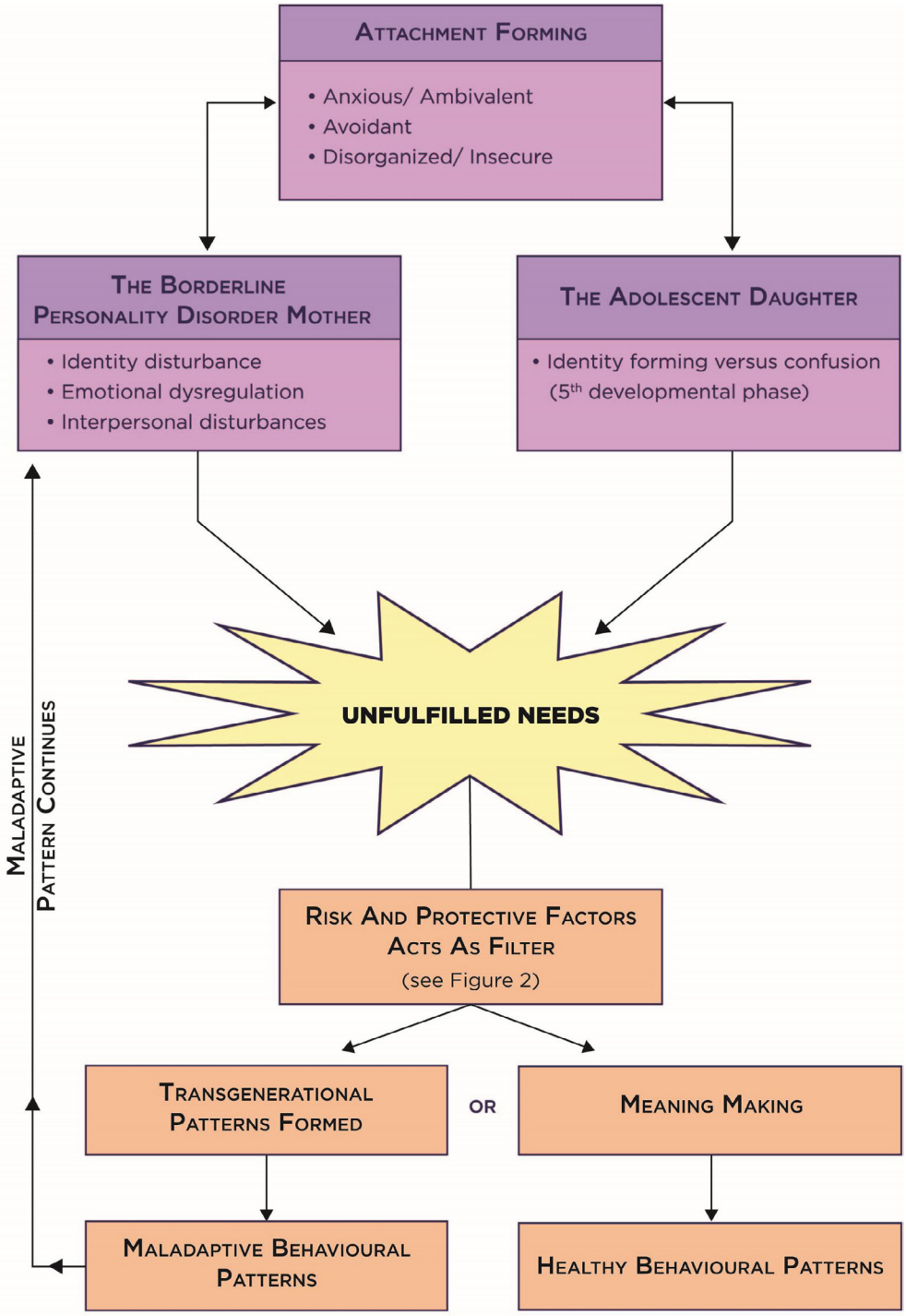
The major objective of this study was to investigate the risk factors for developing a transgenerational pattern and determine what elements will protect the adolescent daughter from this repeated pattern. What will assist the child in reaching a point of existential meaning-making within this transgenerational context? Understanding the philosophical perspectives on meaning-making can inform therapeutic approaches that address the transgenerational patterns in families affected by BPD, and help individuals with BPD develop the skills to regulate their emotions, improve their relationships, and find purpose and meaning in life.

The text explores the philosophical perspective of existentialism, which emphasizes the importance of finding meaning and purpose in one's life. This perspective can help individuals affected by BPD to explore their own values and goals and the patient can be assisted therapeutically to explore what gives meaning to their lives, and to help them to set goals that align with their values. The narrative perspective can also contribute towards meaning-making, as it focuses on the importance of creating new stories that challenge negative self-narratives, and can be helpful in addressing transgenerational patterns of behaviour. By exploring and rewriting family narratives, individuals can gain a better understanding of their own experiences and develop new ways of interacting with their family members. Cognitive-behavioural therapy (CBT) can be informed by the philosophy of empiricism, which emphasizes the importance of testing one's beliefs through experience. In CBT, patients can learn to challenge their negative beliefs and test them through behavioural experiments, which can help them to develop more adaptive ways of thinking and behaving. Furthermore, a mentalisation-based treatment plan to make sense of our thoughts, wishes, feelings and beliefs was addressed. Being an integrative form of psychotherapy, it brings together aspects of CBT, psychodynamic, systemic and ecological approaches. Finally, a psychodynamic perspective can also inform therapeutic approaches for transgenerational patterns in families affected by BPD. Psychodynamic therapy focuses on the importance of exploring unconscious thoughts and feelings, and can help individuals to better understand the underlying motivations behind their behaviour. By gaining insight into these unconscious patterns, individuals can develop greater self-awareness and make more intentional choices in their interpersonal relationships. Most therapeutic approaches emphasize the importance of early intervention that could prevent adolescents from embarking on a maladaptive developmental journey.

This qualitative study is situated within the context of the hermeneutic phenomenological worldview and eleven adolescents were selected for participation. The selection process took place according to the principles of interpretative phenomenological analysis (IPA), which recommends that participants be selected due to their insight into the investigated phenomenon. The recruitment process strived to find a sample that met the homogeneity criteria. The multilayered contexts of research participants, such as their social, historical, economic, and cultural circumstances, are acknowledged by IPA to potentially affect the participant's meaning-making of their experiences. The participants were interviewed by means of semi-structured interviews on two occasions, six months apart. Numerous purposive sampling

designs are available, but this study focused on selecting homogeneous cases to reduce variation and simplify analysis. The goal was to identify and select a group of participants who meet pre-established conditions and, therefore, can be regarded as suitable participants. The researcher purposively sought adolescent daughters with mothers who were diagnosed with BPD. Four themes were deducted from the interviews: (1) experiencing complex interpersonal dynamics, (2) experiencing emotional dysregulation, (3) not managing the system, and (4) having positive expectancies. The importance of trustworthiness and reflexivity were noted.

The study added to the international literature on this topic and emphasizes the importance of psycho-therapeutic interventions as a protective measure for developing a transgenerational pattern. The philosophical perspectives on meaning-making inform the therapeutic approaches of existentialism, narrative therapy, CBT, mentalization-based therapy as well as a psychodynamic perspective and address the transgenerational patterns in families affected by BPD.



Higgins, Wendy (Macquarie University), David Kaplan (Macquarie University), Eliane Deschrijver (Macquarie University), Alexander Gillett (Macquarie University) and Robert Ross (Macquarie University)

Why everyone should be worried about measurement in psychology

Without satisfactory evidence that psychological measurements quantify what they are intended to measure (i.e., they generate "valid" measurements), measurements cannot be interpreted. Thus, it is of considerable concern that recent surveys of measurement validity reporting practices in the social and psychological sciences indicate that validity evidence is frequently unreported, raising concerns about a "measurement crisis". We report results of our ongoing scoping review of validity evidence for the Reading the Mind in the Eyes Test of social cognition, which demonstrate that this measurement crisis is serious. Despite this test having been administered in over a thousand published empirical studies, we discovered that validity evidence infrequently reported and often weak. We also describe two broader philosophical issues with measurement validation that emerged in our scoping review: ill-defined constructs and the treatment of validity as a property of tests rather than test scores. We conclude by suggesting that the lack of critical engagement with validity, treatment of validity as a property of tests rather than test scores, and ill-defined constructs identified in this scoping review are likely relevant to psychological research beyond the RMET literature.

Hochman, Catherine (UCLA)

The Hierarchy of Selves in Perception

The representation of self is deeply connected to many areas of philosophical inquiry, including self-knowledge, self-consciousness, and agency. For this reason, a great deal of work has gone into analyzing de se thoughts – thoughts that include a constituent representation of self (Evans, 1982; Kaplan, 1989; Lewis, 1979; Perry, 1979). But given the central role of the representation of self in our mental lives, it is critical that we also investigate its origins in low-level cognition.

Perception provides fertile ground to examine the origins of the representation of self. The tight connection between perception and one's self, between what is perceived and who is perceiving, prompts the question of whether and how the self is accounted for in perceptual representation. A common philosophical view, which I call Implicitness, holds that the self remains implicit in perceptual representation (Campbell, 1994; Ismael, 2012; Musholt, 2015; Recanati, 2007, 2012). In other words, no perceptual representations include constituent representations of the self as the subject of perception. (Of course, some perceptual representations include a constituent representation of the self as the object of perception, as when, for example, one sees oneself in the mirror.) The motivations for Implicitness are strong, and range from considerations regarding the subject/object distinction, to simplicity, to the epistemic properties of perceptually based judgments. In this paper, I argue against Implicitness and propose a new view – the Nested Frames View – as an alternative. The Nested Frames View holds that there are many explicit representations of self in perception and that these representations form a hierarchy. Information about the self that is implicit in one layer of the hierarchy is explicitly represented in the successive layer. Thus, what we find is that perception involves the use of a complex structure of self representations.

My paper proceeds as follows. After expounding the motivations for Implicitness, I use a prominent view of perceptual representation found in the scientific literature (Ono et al., 2002; Mitson et al., 1976) to rearticulate the notion of a representation of self in perception as the representation of an egocentric reference frame (see Burge 2019, 2022 for a similar idea). An egocentric reference frame is a framework for specifying spatial locations that uses a coordinate system whose origin refers to a point on the perceiver's body and whose axes refer to directions relative to this point.

After reworking our understanding of the representation of self in perception as the representation of an egocentric reference frame, I redescribe Implicitness as holding that no perceptual representations include constituent representations of egocentric reference frames. I then argue that we should reject Implicitness. While Implicitness is usually defended on a priori grounds, I marshal empirical evidence to argue against the view. More specifically, I present two kinds of perceptual experiences that I contend are counterexamples to Implicitness because they seem to involve the explicit representation of an egocentric reference frame. The first case involves the experience of shifting one's gaze. Drawing on Balslev & Miall (2008) and Wang et al. (2007)'s studies on the neural mechanisms underlying gaze shifts in humans and monkeys, I argue that gaze shifts are best explained as resulting from the explicit representation of one's eye-centered reference frame in one's head-centered reference frame. The second counterexample that I consider involves vection, which is the illusory experience of self-motion. I integrate Schwenkler (2014)'s philosophical explanation of vection with a popular psychological explanation (Prothero & Parker, 2003; Riecke, 2011) to reveal that the best computational account of vection involves the explicit representation of one's body-centered reference frame in a room-centered reference frame. I conclude that gaze shifts and vection both seem to constitute counterexamples to Implicitness by involving the explicit representation of an egocentric reference frame.

Having demonstrated how Implicitness admits of counterexamples, I articulate and advance the Nested Frames View. The Nested Frames View accounts for the use of explicit representations of egocentric reference frames in gaze shifts and vection by positing that perception recruits that use of a complex hierarchy of egocentric reference frames. Put roughly, the idea is that the reference frames used in perception are like Russian dolls. Just as one Russian doll is nested within another, which itself is nested within a third, and so on, so too with reference frames: one reference frame is nested within another, which itself is nested within a third, and so on. One frame is nested within another when the origin and axes of the former are explicitly represented in the latter. As before, I take the notions of an egocentric reference frame and a representation of self to be interchangeable.

Once we redescribe the use of egocentric reference frames in terms of the use of representations of self, we find that according to the Nested Frames View, perception uses a hierarchy of explicit representations of self. The hierarchy is such that information about the self flips between being implicit and explicit in each successive stage of perceptual processing.

I end by revealing how this hierarchy of representations of self seems intimately tied to agency. I explain that each representation of self that is added to the hierarchical structure marks a creature's capacity to track a certain set of movements. Because the capacity to track movements is critical to the performance of successful actions, with each additional representation of self, a creature's capacity to successfully act on its desires, beliefs, and intentions increases. In other words, it seems that the greater the number of nested representations of self that a creature uses in perception, the greater its agency. In this way, the Nested Frames View not only deepens our understanding of the use of the self representation in perception, but also provides a new axis with which to measure and analyze agency.

Hofmann, Julian (University of Zurich, Department of Philosophy), Pablo Hubacher Haerle (University of Cambridge, Faculty of Philosophy) and Anke Maatz (University of Zurich, Department of Psychiatry, Psychotherapy and Psychosomatics)

What's the Linguistic Meaning of Delusional Utterances? Speech Act Theory as a Tool for Understanding Delusions

Delusions have traditionally been considered the hallmark of mental illness, and their conception, diagnosis and treatment raise many of the fundamental conceptual and practical questions of psychopathology. One of these fundamental questions is whether delusions are understandable. In this paper, we propose to consider the question of understandability of delusions from a philosophy of language perspective. For this purpose, we frame the question of how delusions can be understood

as a question about the meaning of delusional utterances. Accordingly, we ask: "what meaning(s) can delusional utterances possibly have?" We argue that in the current literature, there is a standard approach to the meaning of delusional utterances, namely the descriptive account which assumes that a delusional utterance "p" means that p is the case. Drawing on Speech Act Theory, we argue that solely relying on the descriptive account disregards essential ways of how linguistic meaning is constituted. Further, we show that Speech Act Theory can prove a helpful addition to the theoretical and clinical "toolbox" used for attempting to understand delusional utterances. This, we believe, may address some of the theoretical and clinical shortcomings of using only the currently predominant descriptive account.

Hornett, Will (University of Cambridge)

Perceptual Capacities and the Mosaic of Sensations

Perceptual experience, at any one time, is incredibly rich with content. I can currently see hundreds of bright, coloured objects, about twelve people in a bustling café, and the spring sunlight falling on the white flowering cherry tree outside. Despite the diversity present in my visual experience, there is a peculiar unity. It is presented to me all in one. My experience seems to be an experience with a number of aspects, parts, or experiential properties, somehow unified. In this talk, I will argue that Schellenberg's popular (2018) attempt to explain perceptual consciousness in terms of perceptual capacities fails to account for the unity of perception. More surprisingly, I will argue that it fails for the same reason that Empiricist analyses of perceptual experience (e.g. Lewis, 1966) in terms of a mosaic of sensations fails. It is surprising because Schellenberg's capacity-based view is supposed to herald an avowedly Aristotelian turn, in appearance utterly different from the Empiricist tradition. First, I will show that Schellenberg is committed to a mosaic of sensations (§1). Then, I will argue that the mosaic of sensations is both incoherent (§2) and makes the unity of perception impossible (§3). Finally, I end with a broad lesson for capacity-based approaches to perception (§4).

§1 – Schellenberg's Mosaic

Schellenberg has argued that discrimination is central both to perceptual experience and epistemology, we can explain it by positing a great variety of low-level capacities to perceptually discriminate things. Schellenberg individuates capacities by what they function to discriminate, and claims that perceptual experiences are manifestations of these capacities. Therefore, you have one capacity for every colour, sound, and shape you can perceptually discriminate. There are two core claims:

1) Every perceptual experience is the manifestation of a capacity to discriminate something

2) For everything one can discriminate, there is a distinct capacity

These jointly entail:

3) If two discriminatory capacities are manifested, there are two perceptual experiences

This commits Schellenberg to a mosaic of sensations. Think about your experience of the colours you can see now. Every portion of your visual field has an associated colour which you see. Each colour is slightly different, and you are employing as many colour capacities as there are visible colours. Therefore, you have as many colour experiences as there are colours visible to you. The colours' location in the visual field can be mapped to a 2D mosaic image, and since the experiences are one-one correlated with the colours, so can the experiences. So, there is a mosaic of simple experiences; your visual experience is a spatially contiguous map of mere discriminations of colour.

§2 – Incoherence

This is impossible; a mosaic of colour experiences is incoherent. To see why, imagine seeing a homogenous red patch, our model for the mosaic's elements. Now, we can obviously discriminate the left part from the right part. The patch has a spatial structure meaning it, itself, is segmented into smaller patches which can, themselves, be discriminated. Therefore, experiencing it means deploying the same red-discriminating capacity for each discriminable part. But, cutting the left part off and looking at that, different spatial

portions of that patch are still discriminable. There cannot be a visible red patch which we cannot do this with. Since the idea of a simple colour experience is one to which this procedure is inapplicable, our model cannot be a patch but a point. But a coloured point is incoherent, since colours are properties of surfaces, and points cannot have surfaces. Therefore, Schellenberg's view entails that there is a mosaic of simple colour experiences, but this is false, there is nothing for there to be a mosaic of.

§3 – Unity

Since the capacity to discriminate red is distinct from the capacity to discriminate squares, (3) entails that if you see a red square, you have two experiences: one of the colour and one of the shape. Two? How is that consistent with the datum that vision seems to present you with an experience of a square? Schellenberg must reject the existence of this 'complex' experience, forcing her to accept either (a) that mere co-consciousness is sufficient for the apparent unity of perceptual experience, or (b) that the unity is post-perceptual, occurring in judgement. I will argue that neither is plausible. The first fails because the occurrence of two experiences is insufficient for an experience of their co-occurrence, which is what we would need (Hurley, 1998). The second raises the puzzle of what grounds a subject could have to judge that two experiences should be thought to be unified if there is nothing in the experiences to justify doing so (Merleau-Ponty, 2012).

In response, Schellenberg may posit a capacity to discriminate red squares, the employment of which is an experience of a red square. But adding this complex capacity doesn't help unless she also subtracts the simple ones. With three capacities, seeing a red square would involve three experiences: one of the colour, one of the shape, and one of the red square, inviting the question how they are unified. The complex capacities must replace the simple one. However, since we can still discriminate colours and shapes, this means rejecting (2), the principle individuating capacities. This blows wide open the question how to individuate capacities. Moreover, since (2) is false, (1), if true, cannot be interpreted as claiming that every perceptual experience is the manifestation of a capacity individuated by its discriminatory function. Therefore, the foundation of Schellenberg's view fails, putting her account of perceptual knowledge, content, and consciousness in question.

§4 – Lessons

The overarching lesson is that capacity-based views of perception must avoid reflecting Empiricist assumptions about experience in their metaphysics of capacities. Instead of picturing capacities as atomic, fine-grained items, we need to think of them more coarsely, not individuated in terms of particular features or objects of experience, but, I will briefly suggest, in terms of certain modes of access to the world associated with the senses.

Isern-Mas, Carme (Univeristy of Granada) and Ivar R. Hannikainen (University of Granada)

The lay concept of self-deception

Which criteria do we use to determine whether something counts as self-deception? In this paper we propose an empirical study to assess the contribution that some controversial features have on the concept of self-deception.

According to the intentionalist view of self-deception, two necessary conditions for self-deception are the agent's intention to deceive themselves, and their having an unwelcome, true belief that they aim to defeat (Demos, 1960; Rorty, 1972). The motivationalist view rejects both claims (Mele, 2001). According to this view, self-deception does not need intention, nor a first true belief to modify. Rather, self-deception is the result of a motivationally biased process to gather and assess evidence.

Parallelly, both motivationalists and intentionalists hold different views on whether self-deception should be characterized as a success condition. Whereas some have argued that self-deception requires acquiring, or at least retaining, a false belief (Mele, 2001, 2010), others disagree, and argue that self-deception might sometimes be a failed process (Audi, 1982; Funkhouser, 2005; Gendler, 2007). In ordinary cases of self-

deception, people do not always manage to acquire, or maintain, a false belief. Therefore, although Mele (2010) has empirically proved that cases in which the agent acquires the false belief are judged as instances of self-deception, attempted self-deception might also count as self-deception on the ordinary usage.

The debate between motivationalists and intentionalists concerns a set of features whose weight in the concept of self-deception is debated. We present two experimental studies testing both views. According to the intentionalists, self-deception requires the presence of an antecedent belief which is contrary to the acquired one, and the intention of the agent to deceive themselves. On the other hand, according to the motivationalists, self-deception only requires the motivation of the agent to believe a proposition. Parallelly, both motivationalists and intentionalists hold different views on whether self-deception should be successful to count as such.

Experiment 1

Study 1 looked at the intentionalist view of self-deception. According to this view, two necessary conditions for self-deception are the agent's intention to deceive themselves, and their having an unwelcome, true belief that they aim to defeat. By contrast, according to motivationalist views, in the garden variety of cases of self-deception the agent does not have the intention to deceive themselves, nor do they need to have a triggering true belief.

Participants (n=160) completed an online survey with a 2 (Intent: present vs. absent) x2 (antecedent belief: present vs. absent) x2 (success: present vs. absent) between-subjects design with an additional within-subjects random effect (scenarios: Beth, Betty, Don, and Sid). In a factorial ANOVA, we examined the effects of intent, prior belief and belief change on perceptions of self-deception. There were significant main effects of intent, $F(1,152) = 4.56$, $p = .034$, $\eta 2 = .008$, antecedent belief, $F(1,152) = 7.56$, $p = .007$, $\eta 2 = .013$, and success, $F(1,152) = 24.45$, $p < .001$, $\eta 2 = .043$, and no significant interactions. Figure 1 displays mean self-deception ratings for each condition (see annex).

Results in Study 1 suggest that the three target features of the intentionalist view have some weight in people's judgments of self-deception. Success turned out to be the most important factor. Antecedent was the second most important factor, followed by intent, which had the weakest effect on judgments on self-deception. On the other hand, unsuccessful cases are still considered cases of self-deception, although less so than successful ones. Therefore, although people do not see success as a necessary condition for self-deception, they see it as a sufficient condition, and their judgments are largely influenced by it.

Experiment 2

Study 2 looked at the motivationalist view of self-deception. Self-deception, according to the motivationalist view, requires a non-deviant motivational cause for acquiring a false belief. By contrast, according to the intentionalists, a motivation is not sufficient because, if that was the case, self-deception would not be different from biases, and precipitated judgments. Therefore, our manipulation of the cause of self-deception included motivated biases, unmotivated biases, and precipitated judgments.

Participants (n=300) completed an online survey with a 3 (Cause: motive, bias, and error) x2 (success: present vs. absent) between-subjects design with an additional within-subjects random effect (scenarios: Beth, Betty, Don, and Sid). In a factorial ANOVA, we examined the effects of cause, and success on perceptions of self-deception. There were significant main effects of cause, $F(2,298) = 11.5868$, $p < .001$, $\eta 2 = .020$, and success, $F(1,298) = 55.3545$, $p < .001$, $\eta 2 = .047$, and no significant interactions. Figure 2 displays mean self-deception ratings for each condition (see annex).

Results in Study 2 suggest that success is sufficient for people's judgments of self-deception, and that the causal role of bias, either motivated or unmotivated, is a necessary condition for self-deception. Consequently, results in study 2 suggest that success, and bias have some weight in people's judgments of self-deception.

Conclusion

Participants in studies 1 and 2 viewed both motivated and unmotivated biases as necessary for self-deception and did not require intention, nor antecedent belief, as necessary conditions. Furthermore, contrary to what intentionalists argue against the motivationalist view of self-deception, people seem to consider bias as a plausible cause of self-deception.

As for the discussion on whether self-deception is a success state or merely a process, people's judgments of self-deception do not depend on the agents' succeeding in acquiring the new belief. Unsuccessful cases

are still considered cases of self-deception, although less so than successful ones. Therefore, although people do not see success as a necessary condition for self-deception, they see it as a sufficient condition, and their judgments are largely influenced by it.

Junker, Frederik Tollerup (University of Copenhagen)

Is the Wandering Mind a Planning Mind?

There is growing evidence that mind-wandering is involved in processes related to planning, practical reasoning, and episodic past and future thinking (Baird et al.; 2011; Ruby et al., 2013; Smallwood & Schooler, 2015). According to recent proposals, mind-wandering is a form of mental exploration (as opposed to an exploitative behavior) that agents can switch to in order to explore new and potentially better action opportunities (Sripada, 2018) or more rewarding goals (Shepherd, 2019) when the value of exploiting known opportunities or current goals is expected to be low. From an action-theoretic perspective, one possible interpretation of this explorative function is that mind-wandering involves reconsideration of prior intentions. Combined with the finding that we spend up to half of our waking hours mind-wandering (Killingsworth & Gilbert, 2010), this raises the possibility that reconsideration is a regular occurrence. Yet this would contradict a central assumption of the influential Planning Theory of Intention (henceforth, PTI) according to which intentions remain relatively stable over time and reconsideration happens infrequently (Bratman, 1987; Holton, 2009). Fortunately, there are other plausible ways of interpreting explorative mind-wandering that avoid this result. I conclude by providing an account of the role of mind-wandering in planning and practical reasoning that is compatible with planning agency.

According to PTI, intentions form partial plans of action that play fundamental roles in practical reasoning and help coordinate our projects over time and with other agents. By committing ourselves to action in advance we are able to make rational decisions in situations where we have too little time to deliberate, or it is too costly to do so. If this picture is correct, it helps explain how limited planning agents can make the best possible use of finite time and limited cognitive resources. But for it to work, the agent's prior intentions must remain relatively stable over time: The agent should not regularly reconsider her prior intentions because if she did, she would have little reason to commit to them in advance. If explorative mind-wandering implies extensive reconsideration this would pose a challenge to the stability assumption of PTI. However, if explorative mind-wandering can be accommodated within PTI, this could lead to new insight into the forms and mechanisms of planning and practical reasoning.

Since during mind-wandering we tend to explore alternative options to those we have previously committed ourselves to, we might interpret explorative mind-wandering as leading to different forms of reconsideration. Usually, our minds wander unintentionally. Any reconsideration during such episodes would be nonreflective reconsideration, i.e., the product of certain habits or dispositions rather than explicit deliberation about whether to reconsider. When, during unintentional mind-wandering, we think of options incompatible with our prior intentions, this could potentially lead us to reconsider whether to stick with our prior intentions. Studies suggest that mind-wandering is also sometimes done intentionally (Seli et al., 2016). We might imagine an agent intentionally letting her mind wander in the hopes of thinking of new options to replace her current intentions. This might be seen as a case of deliberative reconsideration. However, such cases seem to imply the awareness that mind-wandering can be intentionally employed in this way and since mind-wandering is generally considered a relatively passive and purposeless activity, agents probably rarely let their mind wander in this way. Finally, we might imagine cases of policy-based reconsideration where an agent forms a general policy to let her mind wander in the hopes of generating new relevant information, e.g., when current strategies have proved unsuccessful or showed diminishing returns for some time.

Thus, if mind-wandering does lead to reconsideration, it would mostly be in the form of either nonreflective or policy-based reconsideration. However, given the advantages to cognitively limited agents (like us) of forming plans ahead of time and sticking to them, it is unlikely that mind-wandering would have evolved in a way that fundamentally undermines these advantages. More plausibly, the dispositions, habits, and policies that might trigger reconsideration during mind-wandering are limited in frequency and scope (e.g., to

inessential or unpromising intentions) so that they do not generally undermine the stability of our intentions.

In further support of this account, explorative mind-wandering might have other planning-related functions that do not lead to reconsideration. First, mind-wandering might help fill out partial plans by exploring relevant means, preliminary steps, and more specific courses of action. Mind-wandering might enable the consideration of a broader set of possibilities than goal-directed thinking and if this occasionally generates better strategies than would otherwise have been considered, this could justify the time and resources spent mind-wandering by a limited planning agent. Second, mind-wandering might modify one's reasons for doing as one intends without resulting in reconsideration of the intention itself. In such cases, the agent does not reopen the question of whether to abide by her prior intentions but instead simply incorporates new considerations into her reasons for doing what she intends to do.

Finally, mind-wandering might interact with practical reasoning. According to some recent accounts of practical reasoning (Arpaly & Schroeder, 2013; Bendaña & Mandelbaum, 2021; Morton, 2017; Shepherd, 2015, 2022), human agency is characterized by limited cognitive capacity, uncertainty about what to do, and conflict between various beliefs, desires, intentions, commitments, obligations etc. While explorative mind-wandering might be a way to make optimal use of a limited cognitive capacity, it might also seem to deepen the fragmentation of the mind by generating further information that potentially conflicts with existing beliefs, desires, intentions etc. How are human agents able to maintain a modicum of agential unity and rational coherence in light of such fragmentation? One plausible suggestion is that the mental actions we engage in during practical reasoning serve as a kind of rational glue for our otherwise conflicted minds (Shepherd, 2022). On this picture, mind-wandering could be construed as a source of new practically relevant information (e.g., action possibilities or credence updates) which becomes part of the disjointed set of mental states on which the agent enforces a degree of rational coherence through practical reasoning.

Kabrel, Mykyta (University of Tartu)

Extending the mind: the neural basis and phenomenology of introspective insight

Since Tolman's work [10], the importance of spatial navigation has been widely recognised in cognitive sciences. Spatial navigation is the process of finding one's way around in the environment; research has established that it is supported by place cells in the hippocampus and grid cells in the entorhinal cortex [6]. Place cells represent physical space and become active when an animal is in a certain place, while grid cells form a hexagonal grid that covers the space. More recent studies have shown that these cells encode not only space but also other important features like time, sound, taste, learned knowledge, visual categories, social hierarchies, and word meaning [1]. This has led researchers to the theory of mental navigation. It addresses our ability to map cognitive spaces, i.e., navigate and explore our thoughts, memories, and concepts [2-3]. Accordingly, it has been recently hypothesised that insights might occur during mental navigation [1].

In this paper, we spell out and evaluate the rationale behind the hypothesis and expand the framework to the phenomenon of introspective insight as a case study. To begin, we lay out several principles formulated by Aru and colleagues:

- During spatial navigation, neural activity in the hippocampus corresponds to positions in space, while during mental navigation, it corresponds to thoughts, concepts, and memories. Finding a solution during mental navigation is similar to discovering an unexpected reward while navigating through physical space. This could mean discovering a new path or shortcut toward a better understanding of a problem.

- Behavioral timescale synaptic plasticity (BTSP), a rapid form of plasticity, is a relevant candidate to support strong synaptic modifications that occur within the timescale of a single thought process. It is hypothesized to exhibit the following features: (a) quick transformations of the spatial firing properties of the cell, (b) an insight marks a rewarding place in the mental space, (c) due to the rapid plasticity mechanism, insights tend to be remembered better, (d) if the specific inputs from entorhinal cortex and CA3 cells work together properly while navigating mentally, plateau potential arises, provoking the emergence of a new concept field that leads to insight.

- Insight is facilitated by the state of mind-wandering dominated by sharp-wave ripples (SWRs), high-frequency oscillations in the hippocampus. During SWRs, new memory and knowledge combinations are stitched together.

While the literature on insight mainly relies on solving challenging puzzles [5], which require creativity and logical thinking, it leaves the domain of more personal problems underdeveloped. In this paper, we argue that the theory of mental navigation can fruitfully be extended to explain how we resolve inner conflicts. We argue that mental navigation supports introspection and insight, e.g., in psychotherapy or self-analysis.

Notably, introspection is a broad phenomenon: one could, for instance, introspect perception (e.g., qualities of a percept), which is not mental navigation per se. However, when introspection aims to solve a complex psychological problem, it becomes mental navigation. As we argue, the same model can be applied to introspective insight for the following reasons:

- The phenomenology of introspection is highly similar to the concept of mental navigation. During introspection, one needs to unravel a puzzle no less complex than an anagram or a constellation picture. While introspecting on a psychological issue, people encounter impasses in the form of mental blocks, emotions, bodily tensions, breathing restrictions, etc. To achieve a desired state, they need to work hard, to attend to and analyse their thoughts and feelings, i.e., to navigate in the psychological space.

- Spatial metaphors are frequently used to describe our introspective experiences, such as being 'lost in thought,''in a dark place,' or 'at a crossroads.' Our framework explains why such expressions are not mere figures of speech. Recent research indicates that people use space as a source domain for thinking about abstract notions like time, number or emotional valence [7]. This usage applies to introspective thinking, as it utilizes space to describe inner abstract phenomena such as emotion, feeling or memory. Additionally, individuals with extensive experience in introspection commonly employ metaphors such as 'spiritual path' and 'depth of the soul.' This is because they have a subjective sense of distance travelled, with impasses and new pathways expanding their inner domain. Psychoanalysis is often referred to as 'depth psychology' for the same reason.

- We can apply the mechanism of insight to two fundamentally opposite psychotherapeutic approaches, namely CBT and psychoanalysis. In the former case, a therapist asks 'Socratic questions' or proposes to think on the issue 'from the other side.' As mentioned above, neural activity in the hippocampus during mental navigation corresponds to a trajectory of thoughts and memories. Thus, insights operate as sudden discoveries represented as novel concept fields in the hippocampus facilitated by the external input, in our case – a novel thought proposed by the therapist. In turn, the psychoanalytic approach is famous for minimal intervention by a therapist. They apply the 'free associations method,' which uses a kind of mind-wandering consistent with the SWRs. It is intended to encourage sudden insights connecting previously unconnected memories, concepts, etc. It allows one to reorganise unrelated events into coherent narratives (gestalts) that stabilise the emotional state.

- Closing one's eyes or 'looking nowhere' (directing attention inward) during introspection facilitates insight [8-9]. We suggest that we can more easily navigate mentally by reducing external stimuli and leaving only the modality of inner stimuli. This partially explains why meditating or introspecting with closed eyes is generally considered to be more effective. Furthermore, increased alpha-band EEG just before insight suppresses visual distractions, enabling internal association retrieval for solutions, as found [4].

Understanding the role of mental navigation in introspective insight has important implications for mental health, potentially leading to more effective therapeutic approaches and cognitive (neuro) enhancement. If our analysis is correct, similar neural mechanisms support insight during self-exploration and spatial navigation. If so, it should be possible to draw upon the findings from studies using single-cell recordings or fMRI to illuminate the neural mechanisms underpinning introspection.

Keeling, Sophie (Universidad Nacional de Educacion a Distancia)

Responding to Higher-Order Reasons

1 Responding to Higher-Order Reasons

Some reasons are first order: helping others is a reason for volunteering and having fun is a reason

for attending a party. But other reasons are higher-order. They're reasons to exclude certain reasons from consideration, i.e., they're exclusionary reasons (see Raz 1990). And more positively, they can be reasons to respond to reasons. The principles of beneficence might give you reason to not volunteer out of self-interest but instead to do so for the reason that it would help others. And I might have promised my friend to not worry about other people but instead only care about myself for once. This is to say that we can respond to (distinctly higher order) reasons to respond to reasons. But how would this be possible?

Existing work on what it is to respond to reasons has almost exclusively considered the kind of relation involved in responding to first-order reasons but not what it would be to respond to a reason to respond to a reason. Accounting for this is vital for maintaining the plausibility of higher order reasons (HORs) in the first place. A key reason to doubt their existence is that we might question, following Whiting (2017), that responding in this way is even possible.

This paper introduces an account of what I call the 'higher order (HO) basing relation'. Formulating this account performs a double-shift in both developing our picture of higher-order reasons and also defending their existence. While this paper is primarily philosophical, it also poses a challenge to psychologists to develop accounts of the mechanisms at work in HO basing.

Section one introduces HORs and the related notion of responding to a reason to a respond for a reason. As Whiting (2017) writes, for there to be higher order reasons, we must be able to respond to them. Section two canvasses reasons to doubt that we can respond to reasons to respond to reasons. I draw these from the main extant paper on the topic, Whiting (2017). Namely:

• It becomes harder to make sense of HO basing once we emphasise that this relation is meant to be distinct from simply acting/believing for a first order reason. But it's not clear what that would be.

• Normally subjects are creditworthy for acting for good reasons and fail to be creditworthy when they don't. But phiing for a good reason for a pragmatic reason would undercut moral credit – which suggests that HORs are in fact just lower-order reasons.

• There seem to be important limitations on our ability to respond to reasons to respond to reasons. Whiting (2020) notes the case of financial inducement: it just doesn't seem possible to phi for a reason because of a monetary reward.

• HORs seem to be wrong kind reasons (WKRs), but we can't respond on the basis of WKRs. For example, a HOR is a reason to act for a reason for a reason. But this is compatible with it just being a reason to get yourself to act for that reason by getting yourself to believe that the first-order reasons have been disabled.

2

Section three introduces my account. Suppose that I phi for the reason that p for the reason that r. (E.g., I volunteer for the reason that it would help people for the reason that this is an altruistic reason as opposed to an egoistic one). Just as we might think that acting for a reason requires taking the consideration to be a good reason, HO basing requires taking the consideration to be an appropriate kind of good reason, and for the higher order reason to play a causal role in making it such that you act for the relevant lower-order reason. On this picture, HORs are reasons for believing that p is an appropriate kind of reason. When we phi for the reason that p for the reason that r, we are disposed to believe that p is an appropriate kind of reason on the basis of r. And first-order reasons can be excluded by taking the consideration to be an inappropriate kind of good reason. So, for example, the subject would take helping others to be an appropriate kind of reason for volunteering, and they do so on the basis of principles of beneficence. And they would take self-interest to be an inappropriate kind of reason, even though self-interest is nevertheless a (first-order) normative reason.

I then discuss this account's advantages in light of the earlier objections.

• On this picture, responding to a HOR involves a distinctive kind of relation and confers a distinctive kind of normative status onto the lower-order reason and indeed the resulting action. Responding to the lower-order reason is 'higher-order' rationalised by the subject's lights in virtue of how the subject recognizes that the reason isn't just a good reason, but is the appropriate kind of good reason to rely on in that context.

• I can distinguish between wrong-kind and right-kind HORs. Right-kind HORs are reasons for believing that the lower-order reason is an appropriate kind of reason for phiing.

• In general, considering the role of HORs engenders a more complex account of credit. I see no principled reason why HORs can't make a difference to whether someone is creditworthy. HORs can contribute an important form of meta-rationality that makes a difference to our overall assessment of the person.

• According to my account, HO basing will be subject to similar limitations to our ability to form beliefs. We can't choose our beliefs or believe on the basis of financial inducement. And HO basing is subject to

those limitations because of the role played by forming beliefs (namely, the belief that the lower order reason is of an appropriate kind of reason). As such, we can only respond to HORs that we take to be reasons for believing that p is an appropriate kind of reason. Perhaps there will be occasions when financial inducement does constitute such a reason, but they will be rare.

Kersten, Luke (University of Lisbon) and Elmaire Venter (Ruhr University Bochum)

United at Last: An Integrated Account of Predictive Processing and Embodied Cognition

Predictive processing (PP) and embodied cognition (EC) have emerged as two influential approaches to thinking about and studying the mind and cognition. On the one hand, PP maintains that the mind/brain is a hierarchical prediction machine (Clark, 2013; 2016; Hohwy, 2013, 2016). On the other hand, EC maintains that cognitive and perceptual processes are the result of the on-going and dynamic contributions of the body and world (Chemero, 2009; Wilson & Foglia, 2017). Recently, a number of thinkers have started to wonder how compatible the two approaches might be. The trouble is that while many discussions have focused on the relationship between various elements of PP and EC, such as active inference and cognitive scaffolding, less work has been done on showing how the two approaches might be systematically integrated. To advance discussion, we need to move beyond broad discussions of compatibility and focus on developing a well-worked out proposal for how PP and EC might come together. The current paper looks to provide one such account.

We first outline four desiderata on an integrated account:

1. Status requirement – specifying the type of scientific representation being integrated.
2. Commitment requirement – specifying the account's core commitments.
3. Integration requirement – specifying how the integration works.
4. Insight requirement – specifying the insight provided by the account.

These desiderata not only help to guide and constrain discussion but also provide a metric for gauging the strength of a given proposal, representing four plausible constraints on theorising, ones which emerge naturally out of recent thinking on PP-EC compatibility.

We then propose a new, integrated account of PP and EC, what we call the "Embodied Predictive Inference Framework of Cognition" (or EPIC for short). The EPIC emerges from a trio of interconnected ideas.

The first is that the whole embodied agent encodes the causal structure of the world. This is the idea that an agent's body not only plays a nontrivial role in constraining the contents of its model, but also in learning the causal structure of its surroundings. This is done by obtaining a model through dynamic interactions with the world. An agent needs to intervene on the environment in order to disambiguate alternative predictions. The body plays a crucial role, serving to establish a reliable, readily-available 'laboratory' for intervention (Burr & Jones, 2016). We make use of sensory augmentation devices, such as the feelSpace belt, as an example of how embodiment and spatial perception can become constitutively related.

The second is that PEM is widely realised. This is the idea that the body and brain have an equal role in realising the PEM process. The wide realisation of PEM hinges on the hierarchically nested nature of PP systems. At each level of a PP system, inferences are generated by input encoded at lower levels; these inferences are matched with input from the level below generating prediction error that is consequently minimised. This process occurs at each level of the hierarchy, i.e., across the brain and body. An agent's morphology, physiology, behavioural/cultural patterns, and niche all function to keep prediction error within a certain range (Ramstead et al., 2018). If this is correct, then PEM at the lower levels (the scale of reflexes and motor movements) is necessary for ongoing PEM in the neural domain, and vice versa. One interesting implication of wide realisation is that it shifts the so-called 'evidentiary boundary'. Contrary to Hohwy, the EPIC adopts a more inclusive evidentiary boundary, one which incorporates interpretive dispositions, morphology, and neural architecture (Kersten, 2022b).

The third, and final, idea is that prediction error minimisation (PEM) occurs using a single overarching strategy, what might be called "sensorimotor inference". This is the idea that an agent's ability to generate

inferences about the world is conditioned by its embodied actions, taking the form: If action x is performed, then y would be the most likely cause of the structure of the world. Sensorimotor inference involves making predictions about how embodied action affects the causal structure of the world. It involves confirmatory behaviour, i.e., performing action x to confirm y as the state of the world. According to sensorimotor inference, action and perception are not only tightly coupled, but action is essential for perception. If this is correct, then cognitive processes consist in being sensitive to embodied action. This is not to say that every cognitive process relies on embodied action to the same extent. Rather the role of embodied action depends on the task and problem to be solved. The body and brain share an equal role in cognitive processes and the system dynamically adjusts the contributions of body and brain depending on the task at hand.

Having outlined the EPIC, we then return to the four desiderata. First, the EPIC functions as a framework. It does not look to make particular claims about specific cognitive phenomena but, instead, furnish a set of methodological and theoretical guidelines for how specific phenomena can be investigated and explained. Second, the EPIC has three main commitments: namely, those three previously described. Third, the EPIC involves a form of integration via grafting. The key tenets of PP (i.e., that the system models the causal structure of the world and that the function of the system is PEM) and those of strong EC (i.e., that the body is a constituent of cognition) are systemically combined without one set undermining those of the other. Finally, the EPIC furnish explanatory and theoretical insight by providing a toolkit, both computationally and conceptually, that can be used to explain specific phenomena via implementation in particular models and theories. It enables solutions to problems that its competitors fail to make sense of. The wider takeaway from our discussion is that PP and EC are not only compatible in a broad sense, but that a new, hybrid account can be fashioned from their integration. We hope to have shown how PP and EC can be united at last.

Keshmirian, Anita (LMU Munich University), Babak Hemmatian (Beckman Inst. for Advance Science and Technology, University of Illinois Urbana Champaign), Ulrike Hahn (Birkbeck University) and Stephan Hartmann (LMU Germany)

Human vs. ChatGPT: the perceived causal strength in chains vs. common causes

The causal structure is an essential feature of causal reasoning in artificial intelligence (Pearl, 2000) and human cognition (Rehder, 2014). We might have a different representation of causal Chains than Common Causes. However, all else being equal, would such a difference in structure lead to systematic differences in perceived causal strength?

Bayesian Belief Networks (BBNs) provide a common approach to causal structure representation (Pearl, 1988). BBNs are graphs that depict interdependencies among the probabilities of different variables. The variables interconnect through directed arrows into acyclic structures that indicate their probabilistic dependencies. Bayes Theorem provides a prescriptive framework for evaluating outcomes in such networks, meaning deviation from any of its axioms would lead to demonstrably suboptimal reasoning (see Hartmann, 2021). We focus on two canonical BBN structures: Chains (A→C→B) and Common Cause (A←C→B) networks. The joint probability is equal for the two structures. This "equivalence class" means that any systematic difference in perception of causal strength across them is not due to the networks' overall likelihood. However, more important for our discussion is the notion of conditional independence that leads to this equivalence: the probability of B should not depend on A if we know C. In other words, for a given value of C, the characterization of C→B is normatively the same for the Chain and Common Cause BBNs (Figure 1).

However, humans systematically violate independence assumptions in their causal judgments (cf., Rehder, 2014; Park & Sloman, 2013). Recently, the scope of a cause, i.e., the number of distinct effects generated by it, has been studied as a source of perceived causal strength not prescribed by Bayesian theory (Sussman & Oppenheimer, 2020; Stephan et al., 2023). In Chain A→C→B, C's scope is one, less than its scope in the Common Cause A←C→B, which is two. Sussman and Oppenheimer (2020) argued that,

depending on the valence of a target effect (B in this case). For positive results ("boons"), the greater scope shows causal strength. When effects are harmful ("banes"), greater scope leads to lower strength instead. Like most prior research, Sussman and Oppenheimer (2020) restricted their comparison to a single structure type: Common Cause networks, with direct causation as the baseline.

Stephan et al. (2023) criticized the validity of this "Bane-Boon effect". When scenarios were abstract enough to eliminate prior domain beliefs in participants, the effect of scope was uniform across positive and negative target effects. Instead, Stephan et al. (2023) found an additional structural influence on perceived power in Common Cause networks, one they called "the dilution effect" and explained using Power Theory (Cheng, 1997): in causes with greater scopes, a singular "source" of causal power is seen as distributed and "diluted" across the multiple effects.

As it stands, the results of the two papers are incompatible. We directly compare the two mutually exclusive accounts by controlling for valence and varying prior domain knowledge across causal scenarios in human participants (N=300; pre-registration, materials, data, and analysis at https://osf.io/qaydt). We also compare the result of human intuitions to LLMs (i.e., GPT3, n= 60), as many violations of normative criteria in BNNS may have foundations in human language. Prior studies show suboptimal reasoning in causal contexts in large language models (Binz & Schulz, 2023). Since ChatGPT has been trained with language data, this comparison would inform us about the mechanism at play for human intuitions.

To advance the debate about structural influences on perceived causal strength, we also examine whether subjects' perceptions of mechanisms and spontaneous elaborations of the given structures affect causal strength perceptions. Our choice is inspired by Park and Sloman's (2013) study of independence violations in both types of networks. They showed that people use their knowledge of mechanisms to elaborate on the causal structure they are given: subjects' elaborations of causal structures always revolve around mechanisms and could be changed with mechanism-directed instructions. Structures that involve mechanistic explanations have been found to be prioritized both normatively (Russo & Williamson, 2007) and descriptively (Johnson & Ahn, 2017).

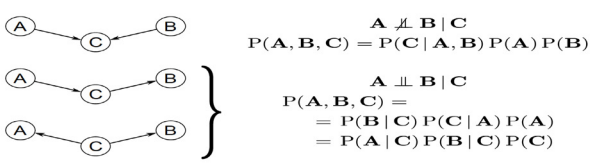
Applying a similar lens to the Chain network without assuming an elaboration of the structures, C may be perceived as the mechanism explaining A's impact on B. If so, the perceived causal strength of C on B in A→C→B would be higher than in a control condition C→B, as the Chain represents a mechanistic explanation absent from the latter case. In A←C→B, on the other hand, the mechanism is vague without elaboration.

If subjects elaborate structures spontaneously by positing connections between A and B, we would expect similar dilutions of C's causal strength in the Chain and Common Cause conditions. If they instead accept the provided networks as the ground truth but base their causal judgments on other structural features like the provision of a mechanism, we would expect the Chain condition to have higher causal power ratings (Figure 2).

Using hierarchical (Bayesian) mixed effect models, we show that - of the possibilities we set out to compare - the subjects' intuition is consistent with a preference for mechanistic chains (Figure 3). This suggests that the perceived boost in causal power in Chains may be due to the perception of the intermediate cause as a reliable mechanism for the network, which is preferred over covariational associations (Ahn et al., 1995). Our models show that chains receive higher ratings regardless of our items' domains, suggesting the robustness of this effect. Our result was replicated in ChatGPT (Figure 4), suggesting that this effect may have foundations in human language. We discuss the implications of our findings for causal representation theories in humans and LLMs.

In addition to our novel finding, we failed to replicate the scope effects in the literature using multiple statistical tests. Despite the apparent negative valence of the effects we used, the greater scope did not produce higher ratings as Sussman and Oppenheimer (2020) theorized or lower ratings, as Stephan et al. (2023) anticipated. This double failure to replicate resolves the discrepancy between the two previous papers' results in an unexpected manner: combined with the mutual exclusivity of their accounts, our experiment casts doubt on the reliability of the Bane-Boon and Dilution effects.

Figure 1. *Joint probability P(A, B, C) is identical in canonical Chain (top) and Common Cause (bottom) Bayesian Belief Networks.*



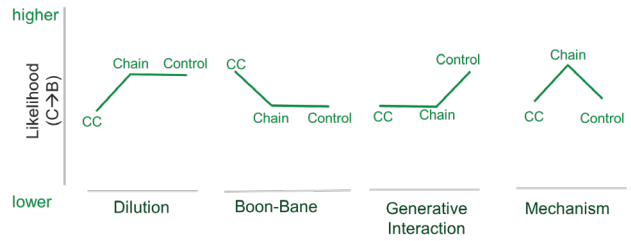


Figure 2. *Theory-inspired pattern predictions for the effect of causal structure on causal power estimates across conditions. The power hypothesis (Stephan et al., 2023) predicts diluted power in Common Cause (CC) conditions due to greater scope, but no difference between Control and Chain where the scope is identical. In contrast, the Boon-Bane hypothesis (Sussman & Oppenheimer, 2020) predicts higher ratings in Common Cause (CC) conditions with negatively valenced target effects. The Interaction account predicts no difference between networks with Preventive links, which are therefore not included in this figure. However, it predicts a lower likelihood in networks with Generative links for both Chain and Common Cause (CC) conditionsIn comparison, a focus on Mechanistic explanation predicts higher power ratings in Chains where the intermediate variable could be perceived as a reliable mechanism.*

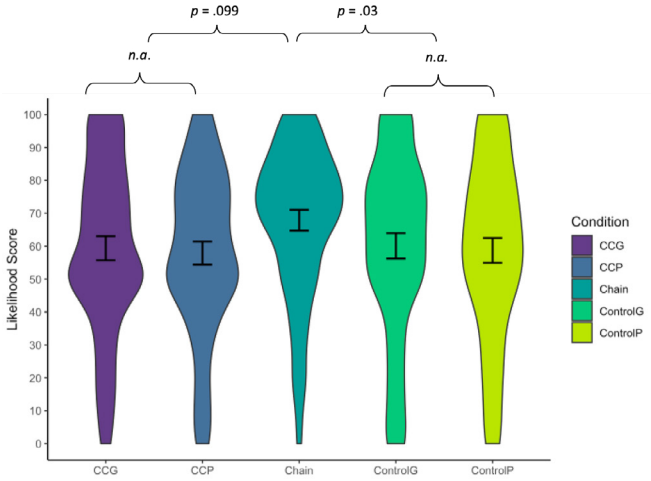


Figure 3. *Distribution of responses over conditions. The likelihood in the Chain Condition (middle row) is significantly higher (i.e., higher likelihood score) than in the rest of the conditions. However, there is no difference between Control (CP; CG) and Common Cause (CCP; CCG) networks. Error bars: 95% confidence interval.*

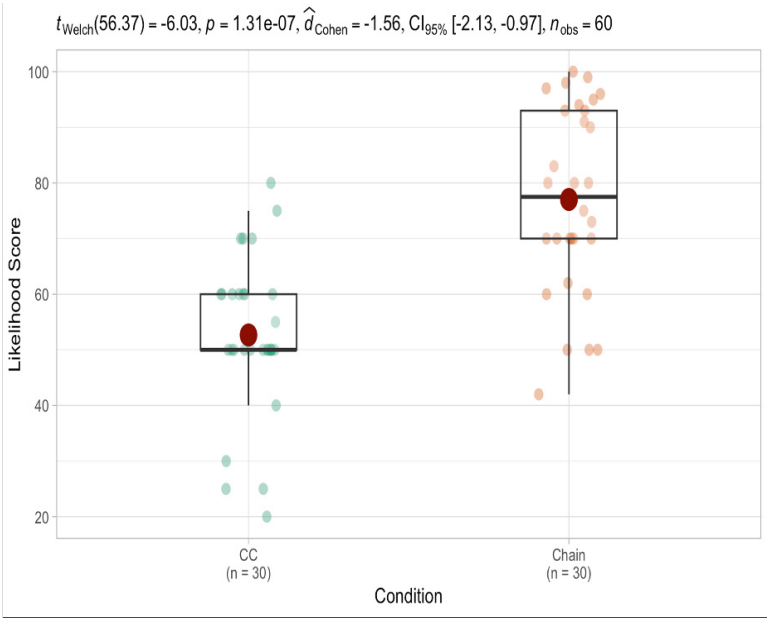


Figure 4. *Distribution of responses over conditions in ChatGPT. The likelihood in the Chain Condition (in orange) is significantly higher (i.e., higher likelihood score) than in the Common Cause Condition (in Green).*

References

Ahn, W. K., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition*, 54(3), 299-352.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104(2), 367–405.

Hartmann, S. (2021). Bayes nets and rationality. In M. Knauff & W. Spohn (Eds.), *The handbook of rationality* (pp. 253–264). MIT Press.

Johnson, J. T., & Drobny, J. (1985). Proximity biases in the attribution of civil liability. *Journal of Personality and Social Psychology*, 48(2), 283.

Johnson, S. G. B., & Ahn, W.-k. (2017). Causal mechanisms. In M. R. Waldmann (Ed.), The Oxford handbook of causal reasoning (pp. 127–146). Oxford University Press.

M. Binz, E. Schulz, Using cognitive psychology to understand GPT-3 (2023). Proc. Natl. Acad. Sci. U.S.A. 120, e2218523120.

Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the Markov property in causal reasoning. Cognitive psychology, 67(4), 186-216.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York, NY: Cambridge University Press.

Rehder, B. (2014). Independence and dependence in human causal reasoning. *Cognitive psychology*, 72, 54–107. *causality*. Stanford: CSLI Publ., pp. 157-172.

Russo, Federica & Williamson, Jon (2007). Interpreting causality in the health sciences. International Studies in the Philosophy of Science 21 (2):157 – 170.

Stephan, S., Engelmann, N., & Waldmann, M. R. (2023). The perceived dilution of causal strength. *Cognitive Psychology*, 140, 101540.

Sussman, A. B., & Oppenheimer, D. M. (2020). The effect of effects on effectiveness: A boon-bane asymmetry. *Cognition*, 199, 104240.

Kliesch, Christian (Department of Developmental Psychology, University of Potsdam)

Infants' Understanding of communicative signals: A 4E-perspective

Humans are an inherently social species. Our lives are deeply intertwined with others from birth, providing the foundation for communication, language, and culture. Many theoretical accounts have argued that such abilities require specific cognitive adaptations that prioritise others as sources of information (Csibra, 2010; Csibra & Gergely, 2009). I propose that these abilities may instead emerge out of humans' prolonged period of helplessness after birth requiring long periods of post natal care. Human infants are born with large brains in bodies that constrain the ability to explore and affect the environment on their own, requiring them to turn to others to affect the environment on their behalf. This unique developmental trajectory provides the foundation of infants' later social understanding.

In order to understand the emergence of species-specific phenotypes, it is crucial to understand the interactions between genes, bodies and environment (Oyama, 2000). There are many examples of these interactions in humans and other animals. Ducklings only learn to recognise species specific calls if they are able to vocalise and hear their own chirps in the egg (Gottlieb, 1997). In toddlers, short arms constrain the number of objects in their field of view, thereby simplifying word-learning (Yu, Smith, Christensen, & Pereira, 2007). As these examples demonstrate, mental processes are grounded in the interaction between the body and the environment. Under the umbrella of 4E-cognition (that cognition is embodied, embedded, enacted, and extended), different philosophical accounts have focussed on the role of the environment and body in cognitive processing (Rowlands, 2009). Taking such a perspective can provide novel insights on the emergence of social and communicative signals in infancy.

Some authors (Csibra, 2010; Csibra & Gergely, 2009) argue for species specific cognitive adaptations that prioritise socially transmitted information over other sources of information, emphasising the discontinuity between the use of social signals in humans and other animals. Other accounts (Heyes, 2012, 2016) suggest that these signals take on meaning through cultural practices. The evidence for a clear demarcation or continuity between human infants' and others animals' use of social signals is mixed. Whilst infants already have perceptual preferences towards social signals, such as eyes and faces by the time they are born (Farroni et al., 2005; Farroni, Menon, & Johnson, 2006; Reid et al., 2017), these preferences are not unique to humans. Infant chimpanzees also prefer direct gaze (Myowa-Yamakoshi, Tomonaga, Tanaka, & Matsuzawa, 2003) and rhesus macaques show

a similar developmental trajectory of gaze following across lifespan to humans (Rosati, Arre, Platt, & Santos, 2016). Despite this, other apes do not use social signals to the extent that humans do (Rosati et al., 2016), and differ in theory of mind (Call & Tomasello, 2008) and communicative abilities (Scott-Phillips, 2015). However, by looking at the specific developmental context, 4E-cognition provides a different explanation based on the embodied, embedded, enacted and extended relationship between infants and caregivers.

Human infants spend a significantly longer time constrained in their motor abilities compared to other species requiring significantly longer periods of postnatal care (Piantadosi & Kidd, 2016). For example, young chimpanzees already start walking and engaging with objects by five months (Potì & Spinozzi, 1994; Yerkes & Tomilin, 1935), about half the age of humans. Such constraints on self-directed movement potentially rebalances the incoming information and learning. For example, monkeys can be trained in mirror self recognition if their movement is restrained (Chang, Fang, Zhang, Poo, & Gong, 2015; Chang, Zhang, Poo, & Gong, 2017). The prolonged period of helplessness and post-natal care shifts infants' direct interactions with the environment to interacting with others. Studies in monkeys and apes suggest that an increase in social interactions between neonates and their caregivers increases social behaviour later in life (Bard, Bakeman, Boysen, & Leavens, 2014; Dettmer et al., 2016; Simpson et al., 2019). In humans, these effects are potentially more profound, priming others as affordances for interaction. For example, eight-month-olds attempt to reach for objects that are out of their reach in the presence of caregivers, but not when their caregivers are not present (Ramenzoni & Liszkowski, 2016).

Children's use of social signals closely follows their motor abilities. Human infants can control their head movements without support and lift their head above the floor only after 3 months of age (Payne & Chang, 2020). In these first months, children in Western societies predominantly see ceilings and the faces of their caregivers (Jayaraman, Fausey, & Smith, 2015; Jayaraman & Smith, 2018). Once children start crawling and walking, they are able to explore the environment on their terms. With the emergence of walking between 9-11 months, children become more interested in distal objects (Karasik, Tamis-LeMonda, & Adolph, 2011), pay less attention to their mothers (Fogel, Dedo, & McEwen, 1992) and rarely lift their heads to look out for others' faces, "because they are too busy playing with toys and running around the room" (Adolph & Hoch, 2019). With upright posture the visual input changes from faces to hands (Fausey, Jayaraman, & Smith, 2016) and pointing emerges out of reaching actions by the age of 12 months (O'Madagain, Kachel, & Strickland, 2019). Parents adjust their engagement to children's emerging abilities (Brand, Shallcross, Sabatos, & Massie, 2007) and provide more learning opportunities to walking children (Kretch et al., 2022).

Looking at the early species-specific environment might help us understand why and how humans, but not other species, develop such a sophisticated understanding of others. The foundations of human social skills may not be exclusively cognitive but emerge from a unique ontogenetic environment that constrains infants' ability to explore their environment on their own. Infants' prolonged dependence on caregivers places them in a niche in which it makes sense to predict others' actions just as well as one's own. 'Others' are integrated into the own body like a tool is integrated by sophisticated users (cf. Gibson, 1986). It is this particular and peculiar environment that provides the foundations of human communication (Pickering & Garrod, 2004, 2014) and social learning (Sebanz, Bekkering, & Knoblich, 2006; Sebanz & Knoblich, 2009).

References

Adolph, K. E., & Hoch, J. E. (2019). Motor development: Embodied, embedded, enculturated, and enabling. *Annual Review of Psychology*, *70*(1), 141–164. https://doi.org/10.1146/annurev-psych-010418-102836

Bard, K. A., Bakeman, R., Boysen, S. T., & Leavens, D. A. (2014). Emotional engagements predict and enhance social cognition in young chimpanzees. *Developmental Science*, *17*(5), 682–696. https://doi.org/10.1111/desc.12145

Brand, R. J., Shallcross, W. L., Sabatos, M. G., & Massie, K. P. (2007). Fine-grained analysis of motionese: Eye gaze, object exchanges, and action units in infant-versus adult-directed action. *Infancy*, *11*(2), 203–214. https://doi.org/10.1111/j.1532-7078.2007.tb00223.x

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192.

Chang, L., Fang, Q., Zhang, S., Poo, M., & Gong, N. (2015). Mirror-induced self-directed behaviors in rhesus monkeys after visual-somatosensory training. *Current Biology*, *25*(2), 212–217. https://doi.org/10.1016/j.cub.2014.11.016

Chang, L., Zhang, S., Poo, M., & Gong, N. (2017). Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images. *Proceedings of the National Academy of Sciences*, *114*(12), 3258–3263. https://doi.org/10.1073/pnas.1620764114

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Brain*, *25*(2), 141–168. https://doi.org/10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*, 148–153. https://doi.org/10.1016/j.tics.2009.01.005
Dettmer, A. M., Kaburu, S. S. K., Simpson, E. A., Paukner, A., Sclafani, V., Byers, K. L., … Ferrari, P. F. (2016). Neonatal face-to-face interactions promote later social behaviour in infant rhesus monkeys. *Nature Communications*, *7*(1). https://doi.org/10.1038/ncomms11940
Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences*, *102*(47), 17245–17250. https://doi.org/10.1073/pnas.0502205102
Farroni, T., Menon, E., & Johnson, M. H. (2006). Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology*, *95*(4), 298–308. https://doi.org/10.1016/j.jecp.2006.08.001
Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005
Fogel, A., Dedo, J. Y., & McEwen, I. (1992). Effect of postural position and reaching on gaze during mother-infant face-to-face interaction. *Infant Behavior and Development*, *15*(2), 231–244. https://doi.org/10.1016/0163-6383(92)80025-p
Gibson, J. J. (1986). *The ecological approach to visual perception*. Psychology Press.
Gottlieb, G. (1997). *Synthesizing nature-nurture: Prenatal roots of instinctive behavior*. New York, NY: Psychology Press.
Heyes, C. M. (2012). Grist and mills: On the cultural origins of cultural learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2181–2191. https://doi.org/10.1098/rstb.2012.0120
Heyes, C. M. (2016). Born pupils? Natural pedagogy and cultural pedagogy. *Perspectives on Psychological Science*, *11*(2), 280–295. https://doi.org/10.1177/1745691615621276
Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *PLOS ONE*, *10*(5), e0123780. https://doi.org/10.1371/journal.pone.0123780
Jayaraman, S., & Smith, L. B. (2018). Faces in early visual environments are persistent not just frequent. *Vision Research*. https://doi.org/10.1016/j.visres.2018.05.005
Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development*, *82*(4), 1199–1209. https://doi.org/10.1111/j.1467-8624.2011.01595.x
Kretch, K. S., Koziol, N. A., Marcinowski, E. C., Kane, A. E., Inamdar, K., Brown, E. D., … Dusing, S. C. (2022). Infant posture and caregiver-provided cognitive opportunities in typically developing infants and infants with motor delay. *Developmental Psychobiology*, *64*(1). https://doi.org/10.1002/dev.22233
Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., & Matsuzawa, T. (2003). Preference for human direct gaze in infant chimpanzees (pan troglodytes). *Cognition*, *89*(2), 113–124. https://doi.org/10.1016/S0010-0277(03)00071-4
O'Madagain, C., Kachel, G., & Strickland, B. (2019). The origin of pointing: Evidence for the touch hypothesis. *Science Advances*, *5*(7), eaav2558. https://doi.org/10.1126/sciadv.aav2558
Oyama, S. (2000). *The ontogeny of information*. Duke University Press. https://doi.org/10.2307/j.ctv1220mm5
Payne, V. G., & Chang, S. H. (2020). Gross motor development. In *The encyclopedia of child and adolescent development* (pp. 1–10). Wiley. https://doi.org/10.1002/9781119171492.wecad023
Piantadosi, S. T., & Kidd, C. (2016). Extraordinary intelligence and the care of infants. *Proceedings of the National Academy of Sciences*, *113*(25), 6874–6879. https://doi.org/10.1073/pnas.1506752113
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169–190. https://doi.org/10.1017/S0140525X04000056
Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, *8*(132). https://doi.org/10.3389/fnhum.2014.00132
Potì, P., & Spinozzi, G. (1994). Early sensorimotor development in chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *108*(1), 93–103. https://doi.org/10.1037/0735-7036.108.1.93
Ramenzoni, V. C., & Liszkowski, U. (2016). The social reach: 8-month-olds reach for unobtainable objects in the presence of another person. *Psychological Science*. https://doi.org/10.1177/0956797616659938
Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The human fetus preferentially engages with face-like visual stimuli. *Current Biology*, *27*(12), 1825–1828.e3. https://doi.org/10.1016/j.cub.2017.05.044
Rosati, A. G., Arre, A. M., Platt, M. L., & Santos, L. R. (2016). Rhesus monkeys show human-like changes in gaze following across the lifespan. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1830), 20160376. https://doi.org/10.1098/rspb.2016.0376
Rowlands, M. (2009). Enactivism and the extended mind. *Topoi*, *28*(1), 53–62. https://doi.org/10.1007/s11245-008-9046-z
Scott-Phillips, T. C. (2015). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Basingstoke, UK: Palgrave Macmillan.
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009
Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x
Simpson, E. A., Sclafani, V., Paukner, A., Kaburu, S. S. K., Suomi, S. J., & Ferrari, P. F. (2019). Handling newborn monkeys

alters later exploratory, cognitive, and social behaviors. *Developmental Cognitive Neuroscience*, *35*, 12–19. https://doi.org/10.1016/j.dcn.2017.07.010
Yerkes, R. M., & Tomilin, M. I. (1935). Mother-infant relations in chimpanzee. *Journal of Comparative Psychology*, *20*(3), 321–359. https://doi.org/10.1037/h0055097
Yu, C., Smith, L. B., Christensen, M., & Pereira, A. (2007). Two views of the world: Active vision in real-world interaction. In *Proceedings of the 29th annual meeting of the cognitive science society*. Mahwah: Erlbaum.

Koi, Polaris (University of Turku)

Nudges alter decision sets

Nudges and other behavioural insights-based steering methods are attractive because, according to the received view, they direct agents to make better choices while 'preserving freedom of choice': for example, Thaler & Sunstein state that they "strive to design policies that maintain or increase freedom of choice" (Thaler and Sunstein 2008, 5).

Whether nudges in fact are freedom-preserving is a topic of lively debate. Proponents argue that nudges are freedom-preserving because the target population is not coerced, and the target population may still behave in an unwanted manner as no option is removed. In other words, when options are not altered or at least not removed, freedom is preserved (Thaler & Sunstein 2008, Schmidt & Engelen 2010, Simkulet 2019, Riley 2017); this conceptualization of freedom draws from 'the cardinality rule' (Pattanaik & Xu 1990). Opponents argue that freedom is not preserved because nudges may 'bypass rationality', harnessing biases instead of appealing to rational choice. In effect, the debate has focused on whether the mechanisms by which nudges influence how we evaluate our options are freedom-preserving. Another claim, namely choice set non-interference, has gone unquestioned.

The present paper offers a novel contribution to this debate, suggesting that choice set non-interference, i.e., the claim that nudges do not alter decision sets, is false.

On strong choice set non-interference, nudges may neither add nor subtract options; on the weak reading, nudges may add options, but may not subtract them. It is plain that nudges sometimes add options. For example, the famous 'Schiphol fly' where a fly-shaped sticker is attached to urinals to improve cleanliness adds the novel option "aim at fly". Nudge defenders may therefore retreat to weak choice set non-interference where only option removal is a threat to freedom. However, as I'll argue, the weak reading is false, as well.

To make this case, we need to distinguish between objective decision sets, such as the number of different brands of cereal at the store, and subjective decision sets, i.e. the decision sets agents in fact deliberate among.

I offer two arguments, both of which establish that even weak choice set non-interference is necessarily false. The first of these is based on insights about the neurocognitive mechanisms of decision set generation, and pertains to subjective decision sets. The second covers both subjective and objective decision sets and is based on a reading of options as packages (Hansson 2018).

In rich environments, agents are faced by too many inputs and action outputs for them to in fact consider all objective possibilities, given human limitations in cognitive processing capacity (Wu 2011). Due to cognitive bandwidth limitations, there is an adaptive preference towards keeping decision sets small as this enables fast and energy-efficient processing; even without this adaptive preference, limitations of working memory capacity delimit the possible range of options in a subjective choice set. Decision-making must therefore be preceded by option generation, a neurocognitive process by which a manageable choice set is produced (Kalis et al. 2013). Neurocognitive heuristics of value ascription, such as proximity and temporal biases and attentional mechanisms, are often thought to influence how we assign value to options; nudges are thought to influence our choices within a given choice set by means of these mechanisms (Thaler & Sunstein 2008). However, these neurocognitive mechanisms also influence option generation by influencing the sampling of information from the environment, a set of processes often referred to with the catchall term 'salience' (Wu forthcoming, Gottlieb 2018). By means of these mechanisms, nudges alter the salience of possible behavioural outputs.

Argument from salience. The more salient a given possible behavioural output is, the more likely it is that this output is included in agents' choice sets. Due to working memory limitations and the adaptive preference for small choice sets, if a given option O is in my choice set, this decreases the likelihood that another option O2 is in this set. Both adding new O and increasing the salience of existing O will thus often exclude other O2….On. Both strong and weak choice set non-interference require that O are not removed; however, the likelihood that some O are in fact removed from subjective choice sets is high. Therefore, choice set non-interference is false.

Some may argue that this is unsatisfactory because the relevant options have to do with objective decision sets, not subjective decision sets. The second argument covers objective decision sets, as well.

Hansson (2018) argues options are not single items, but packages of a set of behaviours and/or considerations: for example, in considering whether to snack on chicken wings or on climate-friendly apples, I consider the sequence of behaviours involved, and the ease, risk, 'fun factor', cost, et cetera of these options. The following argument draws on Hansson's account.

Argument from packages. Nudges influence the ease, novelty, 'fun factor', etc. of a desirable behavioural output. These changes bundle a behavioural output together with a different set of considerations than before. This change in bundling alters the package: for example, instead of 'a scary walk through the park' we get 'a fun walk through the park' as an option. Since options are packages, if nudges alter packages, nudges alter options, in such a way that previous packages are removed and replaced with other packages. Both strong and weak choice set non-interference rule out the substitution of a previous option (e.g., sausage) with another (e.g., tofu). Therefore, choice set non-interference is false.

Thaler & Sunstein, and many others, have taken it for granted that nudges are freedom-preserving due to choice set non-interference. If choice set non-interference is false, however, then this argument for the freedom-preserving nature of nudges does not withstand scrutiny. Whether nudges would necessarily compromise freedom depends crucially on the conception of freedom under scrutiny. Defenders of nudges must revise their conception of freedom, and their arguments in favour of nudges, to incorporate a more nuanced account of the relevant sense of liberty in the context of policy.

Kopsov, Igor (TFMC)

How comparative psychology helps to unravel the mystery behind human motivation

Motivation is studied by a wide range of fragmented disciplines each focusing on specific activities characterized by explicit decision-making criteria. However, humans usually encounter ambiguous circumstances when decisions are made without awareness of commitment. The question of how social settings translate into acts of individual behaviour remains open. We explore how analyzing other social animals' behaviour , e.g. ants, can untangle a blurred picture of socially determined motivation.

Charles Darwin admired ant's "wonderfully diversified instincts, mental powers, and affections" and conceded that "the brain of an ant is one of the most marvellous atoms of matter in the world, perhaps more so than the brain of a man." Ants dedicate their full attention to a specific task at hand, such as gathering food, rearing child ants, housekeeping, and defence, which makes them well suited for analysing behavioural traits.

The behavioural sequence of ants foraging for food includes the following steps: a) chaotic movements; b) when food is found, the route back to the colony is marked with a pheromone trail; c) others follow the pheromone trail; d) subsequent individuals reinforce the trail by similarly marking it using pheromones; e) upon exhaustion of the food source, no further trail markings are made, and the pheromone scent dissipates; f) chaotic movement resumes.

The behavioural sequence of humans looking to satisfy a need includes the following steps: a) seek opportunities, b) upon satisfaction of a specific need, this achievement is marked (visually, verbally, in writing, etc.) describing the trail to the achievement or allowing others to estimate such a trail; c) the trail to need satisfaction is followed by others; d) subsequent individuals reinforce the trail by similarly marking

it through various means; e) Upon exhaustion of the source of satisfaction, no further markings are made, and the trail dissipates; f) the continued search for opportunities is resumed.

The actions of ants following the markers left by other individuals are evident in the behaviour of other species, including ruminant mammals, fish, and birds. Comparing the behaviour of humans to ants shows that humans and other living beings are not too dissimilar. Ants mark paths to success, such as sources of food, by leaving pheromone trails. Humans do it through symbols, written interactions, verbal means, and visual clues. We conceive that trailing the markers of success represents the most generic mechanism behind actions of social nature.

Firstly, humans define a vision of success. Then they link it to a specific set of markers. Most social choices humans make are at least partially based on following success track markers laid out by others. This approach is applied to selecting religious affiliations, places of residence, education, and lifestyle, as well as to more trivial choices such as fashion, music, consumption patterns, and entertainment. Some subcultures generate their style of markers, which can appear obscure to outsiders, for example, particular types of tattoos, social rituals, religious pilgrimages, or celebrations of specific accomplishments. There exist industries dedicated to continuously creating success markers, such as fashion, advertisement, and PR.

Kovačević, Katarina (Central European University, Department of Cognitive Science, Vienna), Francesca Bonalumi (Central European University, Department of Cognitive Science, Vienna) and Christophe Heintz (Central European University, Department of Cognitive Science, Vienna)

You should have checked! The importance of epistemic intentions in ascription of responsibility

People routinely ascribe responsibility for a given outcome. What factors do they take into account during this process? We argue that people ascribe responsibility when they find a counterfactual in which some minimally benevolent intention could have caused a better outcome. We test whether this account accurately predicts how people consider the epistemic states of potentially responsible agents when ascribing responsibility.

We argue that ascription of responsibility is a process that involves mind-reading, causal cognition, and counterfactual thinking. Firstly, mind reading enables us to compute other people's mental states, which is one of the crucial factors in responsibility ascription. Information about whether a person did something harmful intentionally or did not know what the consequences of their action would be, affects the moral judgment of a certain act (Cushman, 2008). In this work, we put a focus on identifying other's epistemic intentions – intentions to take the actions to learn information useful for preventing bad outcomes. For example, checking if the parking spot is reserved for people with disabilities before leaving your car there.

Secondly, causal cognition helps us determine the causal role a person had in producing negative consequences. And finally, counterfactual thinking (Zultan, Gerstenberg, & Lagnado, 2012) enables us to identify the role a certain epistemic state had. If the outcome would have been different if a person knew more, then the (lack of) knowledge was relevant in producing some outcome.

We asked - would protagonists who did not know their actions would have negative consequences be judged responsible, nonetheless? We found that people hold those protagonists responsible if they had an opportunity to acquire information about the potential consequences of their actions, especially if that could have been done easily.

In Study 1 we tested the prediction that ascription of responsibility depends on whether agents had an opportunity to acquire information about the potential consequences of their actions on others. We used three different scenarios and in all of them we varied following three conditions: a) Knowledge - the agent has knowledge about a relevant information (i.e., that their action will bring about a bad outcome); b) Full Ignorance - the agent does not have such knowledge; and c) Willful Ignorance - the agent does not have knowledge about the relevant information, but has the opportunity to take an epistemic action to

acquire such knowledge and willfully does not take it. Participants (N = 151) were recruited via Prolific, they read three stories each implementing a different condition, and rated on a Likert scale to what extent they agreed that the agent was responsible for the bad outcome. In addition, to make sure participants really read the stories, they also answered one attention check question per story, and answers of participants who made a mistake on a certain question were excluded. A Kruskal-Wallis test showed the effect of agents' epistemic states on responsibility ascription ($H(2) = 163$, $p < .001$, $\varepsilon 2 = .383$). Participants rated the agent as less responsible in the Full Ignorance condition compared to the Knowledge condition ($p < .001$) and to the Willful Ignorance condition ($p < .001$); furthermore, participants rated the agent as less responsible in the Willful Ignorance condition than in the Knowledge condition ($p = .01$)

In Study 2 we tested the prediction that ascription of responsibility depends on the (lack of) effort put into acquiring this information. Participants (N = 306) read one story, answered responsibility question, and assessed the difficulty to acquire the information. Mann-Whitney test ($U = 8645$, $p = .002$, $rrb = .194$) revealed that participants ascribed more responsibility to agents who could have easily acquired knowledge and did not (Low effort condition), than to agents for whom that was difficult (High effort condition). In the Study 3 that we are currently conducting we plan to investigate the role of prior probabilities in ascribing responsibility. Do people judge agents who did not acquire the relevant information less in cases when the negative event seemed less likely to happen, e.g. if the exact parking spot was never reserved and a person does not check, expecting it is not reserved this time either?

After conducting the first two studies, we conclude that people weigh costs and benefits for the agent when considering the reasons why they lacked knowledge. In our experimental manipulations, the expected utility of acquiring relevant knowledge in these conditions is informative of the agent's social preferences. If the cost of an epistemic action is low compared to the expected benefit of others, then people infer that the agent who does not perform the epistemic action has low prosocial preferences (Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016). By contrast, if the cost of the epistemic action is high compared to the expected benefit of others, then people cannot make this inference: the agent might have prosocial preferences, which have nonetheless been out-weighted by the cost for the agent.

Our studies suggest that when ascribing responsibility, people consider agents' epistemic intentions and the role they might have had in bringing out the given outcome. This general hypothesis enables making further predictions on how people would judge a person who did not know the bad outcome might result from their actions, but still produced negative consequences. Further psychological experiments could involve varying diverse factors that make an epistemic action worth performing. We varied opportunities to acquire the information and the cost it involved for the agent. Currently, we are investigating whether ascription of responsibility depends on the how likely an epistemic action would make a difference. To have a more detailed perspective on the process of the ascription of responsibility, more research is needed in this field.

Lackey, Nathan (University of Minnesota)

Multiple Model (De)Idealization

Multiple models of the same phenomenon are found throughout the sciences. Most if not all of these models involve some degree of idealization. That is, the models misrepresent the world in some way through inexactness, outright falsehoods, or the omission of some details. There is a rich literature in the philosophy of science devoted to idealization (Potochnik 2017). However, philosophical work on deidealization is sparse. What little work there is in this area has focused on case studies involving a single model or model-type (Knuuttila and Morgan 2019). In this paper, I argue that the process of deidealizing can occur across multiple models over time. To illustrate this, I analyze a case study from the psychology of implicit bias. This novel analysis provides insight into the way that scientists conceptualize phenomena of interest dynamically through multiple models that foreground different aspects of the world. In the first section, I provide an interpretation of the main genres of models for implicit bias. I argue that they are compatible with one another in the sense that their core assumptions are not inconsistent and that they need not be understood as in competition with one another. In the second section, I argue that these

disparate models represent an iterative form of multiple model deidealization (MMD). More recent models emphasize a substantial element that was abstracted away in earlier models. I conclude by demonstrating that deidealization can occur in scientific practice across multiple, compatible models as opposed to just one. This expands upon the framework for analyzing deidealization articulated by (Knuuttila and Morgan 2019) and correlates with the concept of multiple models idealization (MMI) described by (Weisberg 2007).

There are three main types of models for implicit bias in the psychological literature: dual process, propositional, and situationist. Broadly construed, dual process theories hold that there are distinct mechanisms in the mind that dictate behavior in response to stimuli. This explains why people might behave in one way when acting spontaneously under time constraints and another way when they are able to deliberate before acting. These models posit the activation of concepts related in the individual's cognitive architecture (e.g. the concept of "pizza" and "goodness" if one happens to like pizza) (Olson and Fazio 2006). The propositional model states that implicit biases have a structure that is expressible as a proposition (e.g. "pizza is good") (De Houwer 2014). Both models locate the phenomenon of interest in the mind of the individual, as is often the case in psychological theories and experiments. Situationist models break from this pattern in an important sense: these models characterize implicit bias as most accurately located in or distributed across contexts, situations, and structures. "We believe it is more accurate to consider implicit bias as a social phenomenon that passes through the minds of individuals but exists with greater stability in the situations they inhabit…" (Payne, Vuletich, and Lundberg 2017, 236).

I argue that we need not understand these models as being in competition with one another or as inconsistent. It is important to rule out these possibilities because it has been noted that, in some cases, multiple models of the same phenomenon are inconsistent in a way that indicates a problematic lack of understanding (Morrison 2011). My interpretation is that these models represent differences in the values and motivations of the scientists that developed them. This is in line with some observations already present in the literature related to multiple models in the philosophy of science (Weisberg 2007).

In the second section of the paper, I argue that these models, taken together over time, can be understood as a process of deidealization. For example, individualistic models involve a great deal of abstraction: situations and structures are omitted. This would include an individual's home life, where they were educated, the environment in which they work, and the conditions where they make important decisions. In their emphasis on contexts and structures, the scientists generating the new situationist models of implicit bias are in effect deidealizing individualistic models of implicit bias by adding these factors into their models. Although the scientists who developed situationist models may not explicitly see themselves as engaging in a process of deidealization, I argue that this is in fact what has occurred. Importantly, this trend in psychology extends beyond the phenomenon of implicit bias to include accounts of prejudice more generally.

My case study of implicit bias illustrates several conceptual points. First, the philosophical literature on deidealization is in a nascent stage and my analysis is a strategic expansion of the framework provided in the current literature (Knuuttila and Morgan 2019). This expansion correlates notions of MMI with the concept of MMD. Weisberg notes that different kinds of idealizations may arise from the differing motivations held by scientists modeling the same phenomena (Weisberg 2007). I show that this extends to the process of deidealization. Second, Weisberg's account claims that MMI involves models that each have inconsistent fundamental assumptions. I claim that this need not be the case given the example I provide from the science of implicit bias. As a consequence, our understanding of MMI is augmented. All of this provides increased insight into the way that scientists develop and apply models through iterated processes of MMD.

Lee, Ruth (York St John University), Jack Shardlow (University of Edinburgh), Christoph Hoerl (University of Warwick), Patrick A. O'Connor (Queen's University Belfast) and Teresa McCormack (Queen's University Belfast)

Temporal asymmetries and the relative utility of memory and experience

Background

According to philosopher Derek Parfit (1984), upon waking in hospital in a confused state, a person

who is waiting to hear whether they have already had an operation, accompanied by 10 hours of severe pain, or are yet to have a future operation, entailing 1 hour of severe pain, will prefer to learn that they have undergone the past painful operation despite the much greater magnitude of pain that this entails. This thought experiment is intended to demonstrate that people have temporal asymmetries in the value they assign to past versus future experiences (see Hoerl et al., 2022, for interdisciplinary discussion). Empirical studies have since demonstrated that both adults and children do prefer painful experiences to be in the past rather than the future. However, these studies also demonstrated that such preferences are not absolute. Although both adults and children prefer pain to be in the past when considering, for instance, one past versus one future painful event of equal duration (such as a painful injection), they rapidly abandon their preferences when the duration of the past painful event outweighs the duration of the future painful event (Lee et al., 2020). Thus, seemingly contrary to Parfit's thought experiment, past-future hedonic preferences are often abandoned when the quantity of pain under consideration is greater in the past than in the future.

However, the scenarios used by Lee et al. differ from Parfit's in a key respect: Parfit's scenario, unlike those of Lee et al., specifies a mechanism (a post-operative drug) that ensures the painful operation will never be remembered. Because the recollection of past events can itself be aversive or enjoyable, memories might carry their own (dis)utility (Elster & Loewenstein, 1992; Kahneman, 1999; Morewedge, 2015). Thus, Lee et al.'s (2022) subsequent studies directly examined whether past-future hedonic preferences might be affected by the utility people assign to experiential memories. They found that when adults are told that their memory of the pain will never return, they prefer past pain that is tenfold greater than future pain, in line with Parfit's intuition; however, this preference disappears when they are told that, despite their current confusion about events, they will eventually remember the pain. These contrasting findings suggest that memories are treated as goods with their own (dis)utility, such that memories of pain are negatively weighted in overall calculations about the utility afforded by some event, and in this way influence the extent to which preferences for the temporal location of pain are asymmetrical.

The current study

Lee et al.'s findings suggests that (dis)utility is assigned to memories of experiences, but, because participants always faced a choice between pain in terms of where it is located in time (past versus future), the results do not allow straightforward inferences to be made about the relative utility assigned to memories of experiences versus experiences themselves. The current studies directly investigated the relative utility of memories of pain versus experiences of pain by using a scenario in which participants chose between two events that both occurred in the past (Past condition). Participants were asked to express a preference between 10 hours of past pain that will never be remembered, and 1 hour of past pain that will be remembered. By varying the framing of the test question, we also manipulated whether participants were waiting for news or expressing a hypothetical choice (about who they would rather be), to examine whether preferences might be influenced by a sense of agency. Finally, we also used a Future condition in which participants made the same choice (10 hours of unremembered pain versus 1 hour of remembered pain) but both events were located in the future, because we suspected that, when anticipating future events, people will place more weight on experiences themselves rather than their future memories of such experiences.

Findings

In the Future condition, as predicted, there was a very strong preference for 1 hour of future pain that would be remembered over 10 hours of future pain that would be forgotten, regardless of the framing of the test question. By contrast, in the Past condition, participants were at chance when stating a preference between hearing news that they had experienced 1 hour of to-be-remembered pain versus that they had experienced 10 hours of to-be-forgotten pain. Question framing had an effect in the Past condition: when asked to choose who they would rather be, someone who had experienced 1 hour of to-be-remembered pain or someone who had experienced 10 hours of to-be-forgotten pain, they chose the former significantly more often than chance. This may suggest that when there is some sense of (albeit hypothetical) agency over past events they are not represented as 'over'. However, participants nevertheless chose the 1 hour of to-to-be-remembered pain less frequently than in the equivalent Future condition.

Thus, the temporal location of pain appeared to influence the disutility assigned to memories of pain relative to those assigned to painful experiences themselves: when pain is located in the past, this seems to

enhance the importance of its memory. This may be for several reasons, which are not mutually exclusive. First, people may assign less value to the memories of past pain than to the imagination of future pain. Future events elicit more emotion and feel closer than past events (Caruso, 2010; Caruso et al., 2013; Van Boven & Ashworth, 2007); it is possible that this adds to the weight of their disutility as experiences through which to live, and/or the very anticipation of pain may carry disutility. Second, in the Past condition, the fact that the pain is over may render its duration less important; and memories of the pain would, if not for their permanent erasure, now be looming in future prospect.

Work now in progress seeks to establish whether and at which point a further increase in the duration of forgotten pain for those who initially preferred it, or an increase in the duration of remembered pain for those who preferred that scenario, results in a change of preference.

Leising, Daniel (Technische Universität Dresden), Oliver Grenke (Technische Universität Dresden) and Marcos Cramer (Technische Universität Dresden)

An Advanced Tool for Comprehensive Argument Visualization

Preprint is available here: https://psyarxiv.com/dvfq7/

The Visual Argument Structure Tool (VAST; Leising, Grenke & Cramer, 2021) is a highly generic tool that enables comprehensive visualizations of the semantic, conceptual, empirical and reasoning relationships that constitute arguments. It builds on classical propositional logic but also incorporates elements such as "measurement error" from psychometrics.

VAST does capture the difference between mental concepts and the words that may be used to designate them, as well as various ways in which concepts themselves may be related to one another. All of this is distinguished from beliefs as to whether something IS the case and/or OUGHT to be the case. The system also accounts for WHO thinks that something is or should (not) be the case, and for the difference between concepts that do have empirical reference points (i.e., data) and concepts that do not. Any combination of basic system elements may be grouped together and thus form a new element that may have any kind of relationship with other elements. The latter feature enables analyses at different levels of resolution. Despite its comprehensiveness, relatively high parsimony of the system is achieved by using a small set of graphical elements (e.g., boxes, arrows, ovals) with well-defined meanings, and a unified metric of relationship strength.

VAST enables representing argument structures at any level of vagueness vs. precision that is deemed possible and/or necessary. For example, a VAST display may comprise both narrative content and objectively measured data points, and specify how these relate to each other. This feature may make the system particularly useful for attaining greater theoretical specificity in the social sciences, where theories are often rather broad and sometimes ambiguous in regard to their implications and testability.

Majeed, Raamy (University of Manchester)

On How to Develop Emotion Taxonomies

How should we go about developing emotion taxonomies suitable for a science of emotion? Scientific categories are supposed to be "projectable": they must support generalisations required for the scientific practices of induction and explanation (Goodman, 1954). Attempts to provide projectable emotion categories typically classify emotions in terms of a limited set of modules: systems/mechanisms/programs hardwired into our brains by evolution and purpose-built to generate certain coordinated patterns of expressive, physiological, behavioural and (perhaps) phenomenological responses (Griffiths, 1990, 1997,

2007; Charland, 1995; Cosmides and Tooby, 2000; Ekman, 2003; Prinz, 2004). The basic idea is that our emotion categories will be a reflection of our emotion modules, e.g., if there is a fear module, then there is a category fear, and if there is no fear module, then there is no such category.

Attempts to categorise emotions in this manner typically result in emotion categories which overlap with Ekman's (1973, 1999) list of basic emotions. That is, our emotion categories include basic emotions like fear, anger, happiness, sadness, disgust and surprise, but not complex emotions, e.g., jealousy, humiliation and love. The reason for this is that complex emotions don't stem from emotion-specific modules (or even a combination of such modules), and therefore they won't provide us with projectable categories suitable for scientific inquiry. For instance, while we may extrapolate from some instances of the category anger to other instances of the category, we can't do so with complex emotions, such as jealousy. Anger stems from an anger module that ensures that episodes of anger will have certain underlying features in common, whereas there is no jealousy module that guarantees that we can make similar generalisations about the category jealousy by simply observing some of its instances (Griffiths 1997).

While modular emotion taxonomies have been hugely influential, there have also been highly controversial. The main criticism with respect to categorisation is that they misrepresent the diversity of our emotional repertoire. There are actually a series of concerns here. One specific concern is that to categorise emotions in terms of a set of evolved modules is to ignore the sheer variety of emotions we can experience (Nussbaum, 2001; Solomon, 2007; de Sousa 2008). Most of us accept that we experience more than just basic emotions; we experience emotions like jealousy and humiliation. Some philosophers, e.g., de Sousa (2008), go even further and argue that we experience "innumerable" emotions, much more than those for which we have words.

There are, of course, plenty of non-modular conceptions of emotion. Some, e.g., Barrett (2017), argue that emotions are social constructs. Others, e.g., Adolphs and Anderson (2018), argue that emotions are best investigated functionality. de Sousa himself (2008) argues that emotions should be understood using multidimensional emotion scales. While such accounts promise to deliver more inclusive emotion categories, what is missing from them is an explanation of how such categories might also prove projectable. Without an explanation, it appears that we are in the midst of a dilemma. We either have a method of categorising emotions that will ensure projectability at the cost of emotional diversity, or we have methods that can help us capture a more inclusive set of emotional phenomena at the cost of having projectable emotion categories.

In this paper, I explain how a developmental approach to emotion, one which focusses on ontogenetic development (i.e., the development of a train during an individual's lifespan) instead of phylogenetic development (i.e., the development of a train within a species) can help us categorise emotions in a way that respects both their diversity and projectability. The key concept is Karmiloff-Smith's (1992, 1994, 2009, 2015) notion of modularisation, which plays a crucial part in both the Developmental Systems Theory (Griffiths and Gray, 1994; 2004) and Neuroconstructivism (Westermann et al., 2007; Karmiloff-Smith et al., 2018). The basic idea here is that the development of a trait depends not just on an innate, genetic "program", but on "multidirectional" interactions between our genes, brain, cognition and environment (Karmiloff-Smith 2009).

Crucially, such interaction can lead to a process of "modularisation", which result in traits with modular features (e.g., automaticity) without actually being the products of innate modules themselves. Emotional traits that appear modular can be understood in this manner. Some emotions, e.g., episodes of fear, can be quick, automatic and pre-reflective, but this need not be because there is an innate fear module. Rather, episodes of fear can manifest such modular traits through the developmental process of modularisation (Majeed, 2022a, 2022b).

In this paper I argue that modularisation offers us a more inclusive way to think about emotion categorisation. The key question driving this debate is the question of projectability. Simply put, What makes emotion categories projectable? According to Griffiths (1997), what ensures projectability is descent. As he argues, in comparison to classifications based on analogy (shared functions), those based on homology (shared ancestry) are "deep": even when the function has been transformed, there is more convergence in the underlying causal mechanisms (Griffiths 1997). Griffiths's point is that it is by looking at descent that we find the relevant mechanisms, neural or otherwise, which explain the projectability of the category.

As we have seen, this does explain the projectability of our emotion categories, but only by limiting emotion categories to those that stem from evolved emotion-specific modules. An altogether different way to think about projectability is via moduralisation. What ensures projectability is not that our emotions stem from the same evolved mechanism, but rather that they emerge from the same underlying process,

namely modularisation. This proves advantageous in that modularisation is not a process that operates just on certain evolved or basic emotions. Take a complex emotion such as jealousy. Modularisation can explain why certain instances of jealousy can have certain modular traits, e..g, be quick, automatic and pre-reflexive. In this way, modularisation can help deliver a more inclusive list of emotion categories which are also suitable for scientific inquiry.

Matey, Jennifer (Southern Methodist University)

Grief and Meaningfulness

Most would agree that whatever 'the good life' comes down to, it must be to some degree be a life that is judged to be meaningful. But what makes a life meaningful? Traditionally, philosophers fall into two groups in answering this question. Some identify objective criteria. For example, C.S. Lewis claimed that meaningfulness involves appropriately relating to God. Research by F. Martela suggests that lives can be considered meaningful when they contribute to something beyond themselves. But it seems possible to meet either of these criteria and still feel that one's life lacks meaning. An alternative set of subjectivist views ground the meaningfulness of one's life in subjective facts. For example, one might be said to lead a meaningful life if and only if: one's desires and preferences are met, or one accomplishes what one believes to be important, or one has pro-attitudes such as caring, love, satisfaction or other affective states. But even granting that these criteria may count as sufficient conditions for a meaningful life, are any one of them really necessary?

Drawing on phenomenological and epistemological facts concerning the complex experience of grief, this paper makes the case that the subjective experience of meaningfulness is more commonplace than it might generally be thought to be. Grief, I argue, is a special type of evaluative emotional experience that follows the loss of a grounding object such as a spouse, a career, or a child. It is a special type of evaluative emotional experience that provides information about the experience of meaningfulness that was present prior to that loss. I argue that grief is especially suited to offer this insight insofar as the lack of meaning experienced in grief is experienced as contingent; it involves observing a sudden experience of a lack of meaning, and by comparison, an imagined counter-factual state in which the loss has not occurred. It is reasonable to take that counterfactual state to contain a property, meaningfulness, that it now lacks. The argument from grief is made in part by contrasting the dynamic perspective shifting between the actual and counter-factual states in grief with the lack of meaning that is experienced by people in major depression as seeming necessary and inescapable.

Two additional questions arise. What is the source of the meaningfulness that grief exposes and why do we need to experience the loss of meaningfulness in order to discover recognize it? With respect to the first question, according to the view I defend, even the most mundane and routine moments of one's life can be imbued with a kind of meaningfulness that is grounded in, or supervenes on, a facet of the subject's psychology- what I refer to as their 'evaluative stance'. An evaluative stance can be described as a kind of a standing psychological state that disposes a person to find specific things to matter and have value as well as the specific ways of mattering and of being valued that things are taken to have. For example, a job may matter as the means of supporting one's child. A home may have value as the location for time spent with a spouse.

One's evaluative stance depends on particular objects that serve to ground the significance and value that is attributed to the various activities and features that comprise their day-to-day life. These grounding objects need not be limited to people one relates to, but can include a religion, a career, an aspiration, or even a nemesis. What is important is not the kind of thing that the object is, but rather the role that the object plays in how one is disposed to live as well as what values and ways of mattering the features of one's day-to-day life one is disposed to take them to have.

With respect to the second question, my claim is that this type of meaningfulness is an aspect of phenomenal character that, when it is present, we have only tacit rather than focal awareness of. Tacit awareness is a type of awareness that provides context for what is attended to, but what we experience

only tacitly cannot be reported on. This explains why people are not disposed to form the kind of beliefs about the meaningfulness of mundane aspects of their lives that lead them to know that the routine and mundane is in fact experienced as meaningful. So subjects are left to learn about that meaningfulness in a different way, by inference.

Meertens, Nadine (Ludwig-Maximilians-Universität Munich)

A Functional Account of Awareness in Animals

Imagine the scenario in which a lizard approaches a cricket. After sensing its target it moves towards it, and adjusts its path to catch up with the moving cricket. Once the lizard gets close enough to the cricket it promptly eats it. Is this lizard aware of the cricket?

Intuitively it seems most people would be inclined to answer this positively – after all, when your cat does the exact same thing you would be loath to deny it awareness. Our inclination to grant awareness in daily life need not be reflective of if it is actually present. Whether the lizard would have a representation of the cricket is also something that has garnered much discussion (e.g. Millikan, 1989). Yet, what it means for an animal to be aware – as distinct from having representations – is still an open question in the philosophical literature. There is no clear and established definition, or model of awareness. The scope of the concept of awareness is even more confused due to its constant conflation with concepts such as consciousness, subjective experience, sentience. I will take awareness to be different from phenomenal consciousness, and as referring to something more minimal. For most people we feel safe saying they are aware because we can extrapolate from our own experience, talk to them using language, and measure their brain activity. The problem of other minds – when it comes to human animals – is more or less solved. But, what about the lizard?

I argue that a functional account of awareness for animals can provide us with the tools in order to formulate an answer to the questions of awareness. Two key questions can be highlighted when it comes to awareness: (i) what is awareness; (ii) how do we tell if a certain animal is aware? I propose that in order to gain clarity on both (i) and (ii), one must ask a third question, namely: (iii) what is the function of awareness? It is the exploration of the third and second question that can offer some clarity on the first.

To investigate the function awareness serves for certain animals within their experiential niche, a preliminary answer to the first question is needed. The focus is not on consciousness, or it 'being like' something to be in a certain state (Nagel, 1974). I propose instead to start with the operational definition of awareness in terms of privileged access. Awareness is defined as the functional ingredient that allows for certain capacities, or behavioural organisation to take place. Having access to the environment in a way specific to me is what allows me to focus on specific elements relevant to my tasks, engage with them flexibly, and so forth.

In the ideal scenario the end result would be a framework for awareness that is operational (allowing for testing and comparison), but also minimal. The framework also needs to allow for various ways of instantiation. The framework needs to make sense for many different animals, and avoid being formulated in terms that describe only one kind of animal (and anthropomorphism). The function awareness serves within different animals' experiential niche and environment need not be the same, or may present in different ways. Two species might be aware in different ways, and of different contents.

The functional account I propose takes the shape of a standard modus ponens argument: P1. X needs awareness to ϕ

P2. ϕ

Thus X is aware

X is a placeholder for a species of animal, or classes such as mammals. ϕ stands for capabilities, or capacities to engage or behave in certain ways. For example: a lizard might need spatial awareness to behave flexibly in unknown environments. So, if we can show via a number of tasks or tests that the lizard can indeed behave flexibly in a range of new environments then we can say the lizard has spatial awareness. The functional account would allow a researcher to start small and work their way up from there. By investigating and expanding upon this example of the lizard, the lizard can be contrasted with

other animals. This comparison will allow for patterns to emerge that can aid us in revising and expanding the operational definition of awareness that started the investigation.

This relies on the notion that it only makes sense to posit awareness in any being if it is required to explain their capacities and behaviour. Is it the best explanation, or can we explain the capabilities of the animal in purely mechanistic terms – making attributions of awareness superfluous? The idea is that we are looking for ϕs where awareness is a necessary condition (does not need to be sufficient). The success an animal has in engaging in tasks that test for the presence of such ϕ will allow for awareness to be measured.

In order to establish this I will draw on a variety of discussions from the literature on agency in animals (Tomasello, 2022). Several examples of animals will be drawn upon to illustrate possible ϕs, and the function awareness has within the animals' experiential niche. The theoretical underpinnings of the functional account will draw on previous functional accounts of consciousness, such as that by Dennett (1991). The preliminary definition of awareness will be elaborated on, and contrasted with access consciousness as formulated by Block (1995).

In doing this I make the case for a functional account of awareness as a way of making the discussion on awareness in animals more concrete, operational, and minimal. By asking what the adaptive advantage of having awareness might be for various different animals within the animal kingdom, clarity can emerge on the phenomena we are intending to describe. What are the commonalities, what keeps the mongrel concept of awareness together?

Melis, Giacomo (University of Stirling)

Rational belief-formation and reflection in human and non-human animals

Rational agents are capable of forming beliefs in response to reasons. Many philosophers (e.g. Boyle 2012, 2016, 2018, Marcus 2021, 2022, Korsgaard 2009, 2018) advocate a reflective characterization of responsiveness to reasons, according to which a belief may count as rational only when the believing subject understands what her reasons are and can critically evaluate them. Under this characterization, the capacity to reply to "why?" questions is typically seen as necessary to engage in rational thought. Thus, only subjects capable of speaking human languages are taken to be capable of genuine responsiveness to reasons. The view allows that non-human animals enjoy various forms of intelligence, but it denies that they are able to take ownership of their beliefs in the way that is seen distinctive of rational subjects.

I challenge this way of defending human uniqueness with respect to rationality in two ways. First, I argue that the view struggles to make sense of human children's transition to rational agency. Second, I contend that the view is committed to a discredited form of internalism in epistemology, and that appreciating what is wrong with such internalism suggests that we must make room for unreflective responsiveness to reasons.

Here's a quick summary of the first concern. Consider young children who are not yet able to reply to "why?" questions, but who can interact intelligently with other subjects and the environment. For example, they may be able to assess their own epistemic situation in relation to someone else's (e.g. Haigh and Robinson 2009), grasp the difference between informative and non-informative access to objects (e.g. Robinson et al 2008), or have some grasp of the normative structure of conventional games (Rakoczy, Warneken and Tomasello 2008). Are these children rational subjects? According to the letter of the reflective characterization of rationality, the answer would seem to be "no." If so, children's transition to rational subjects is mysterious: how can rationality suddenly appear in a non-rational mind? If, on the other hand, the advocate of the reflective view would like to answer "yes", an account of how the abilities mentioned relate to the capacity to reply "why?" questions is needed. Saying that infants are rational because one day they will be reflective is no less mysterious than saying that at some point in cognitive development the non-rational mind of a subject is transformed into a rational one. Either way, an account of how subjects become reflective that ties with the claim that only subjects capable to reply to "why?" questions are rational is needed.

Here is a summary of the second concern. Since replying to a question like "why do you believe that P?" requires one to individuate one's grounds for believing P, the view under scrutiny has it that genuine

responsiveness to reasons requires one to be able to identify one's reasons. This is, in effect, access-internalism: a theory of epistemic justification according to which one is justified in believing P only if one is in the position to access one's evidence for P on reflection. One problem with access-internalism is that it leads to infinite regress. Consider that one couldn't have a reason for belief unless one had an attitude towards it, be it another belief, a perception, a seeming, and so on. Now, that attitude had better be rational itself, on pain of irrationality of the original belief. Thus, the acquisition of reasons in support of a given attitude must itself be rational. Suppose now that one formed the belief that P on the basis of Q. According to the reflective view, one must be in the position to individuate Q as a reason for P. However, if the acquisition of Q is itself rational, one also must be in the position to individuate some further reason R in support of Q. But if the acquisition of R itself is to be rational, then one also must be in the position to individuate a further reason S in support of R, and so on. Regress ensues (cf. Kornblith 2012, Wright 2012).

The upshot of the regress argument is that, at some stage, one must be rational in forming an attitude without being in the position to identify one's reasons for it. In other words, there must be cases of brute rationality—cases where one can have a rational belief without being in the position to individuate one's reasons and appreciate their strength. Once we acknowledge this, it is a short step to conceive of subjects who are unreflective through and through. Such unreflective subjects would have rational beliefs, but would be in no position to individuate and assess in thought the reasons for any of their beliefs. Many have indicated very young children and some animals as likely candidates (e.g. Kornblith 2021, Graham 2018, Bermudez 2003, 2018, Carruthers 2019). While at present, we can't say which subjects or species may be capable only of unreflective responsiveness to reasons, in the light of the considerations rehearsed above, we must take seriously the possibility that there are some such subjects. If so, we are under pressure to allow for the possibility of genuine responsiveness to reasons in non-human animals. This is an option that some philosophers (e.g. Glock 2019 and Dretzke 2006), have begun to explore, and that sits well with Conee and Feldman's (2001) mentalist-internalism in epistemology.

It is possible that one of the motivations for the view that the only genuine notion of responsiveness to reasons is the reflective one lies in the difficulty to conceive of the transition between unreflective and reflective responsiveness to reasons. Time permitting, I will conclude by outlining a way in which such transition could take place, which does not presuppose possession of linguistic abilities.

References

Bermúdez, J. L. 2003. *Thinking Without words*. Oxford: Oxford University Press.

Bermúdez, J. L. 2018 Can Nonlinguistic Animals Think About Thinking? In: Andrews, K. And Beck, J. (eds.), *The Routledge Handbook of Philosophy of Animal Minds*. Routledge: London and New York, 119-130.

Boyle, M. 2012. Essentially Rational Animals. In: Abel, G. and Conant, J. (eds.), *Rethinking Epistemology*. Berlin: de Grutyer, 395-428.

Boyle, M. 2016. Additive Theories of Rationality: A Critique. *European Journal of Philosophy*, 24 (3): 527-555.

Boyle, M. 2018. A Different Kind of Mind? In: Andrews, K. and Beck, J. (eds.), *The Routledge Handbook of Philosophy of Animal Minds*. Routledge: London and New York, 109-118.

Carruthers, P. 2019. *Human and Animal Minds.* Oxford: Oxford University Press.

Conee, E. and Feldman, R. 2001. Internalism Defended. In: Kornblith, H. (ed.), *Epistemology: Internalism and Externalism*. Oxford: Blackwell, 231-60.

Dretzke, F. 2006. Minimal Rationality. In: Hurley S. and Nudds, M. (eds.), *Rational Animals?* Oxford: Oxford University Press, 107-116.

Glock, H. J. 2019. Agency, Intelligence, and Reasons in Animals. *Philosophy*, 94: 645-671.

Graham, P. 2018. Formulating Reductionism About Testimonial Warrant and the Challenge from Childhood Testimony. *Synthese*, 195, 3013–3033.

Haigh, S. and Robinson, E. 2009. What Children Know About the Source of Their Knowledge Without Reporting It As the Source. *European Journal of Developmental Psychology*, 318-336.

Kornblith, H. 2012. On *Reflection*. Oxford: Oxford University Press.

Kornblith, H. 2021. *Scientific Epistemology: An Introduction*. Oxford: Oxford University Press.

Korsgaard. C. 2009. *Self-Constitution: Agency, Identity and Integrity*. Oxford: Oxford University Press.

Korsgaard, C. 2018. *Fellow Creatures*. Oxford: Oxford University Press.

Marcus, E. 2021. *Belief, Inference and the Self-Conscious Mind*. Oxford: Oxford University Press.

Marcus E. 2022. Anscombe and the Difference that Rationality Makes. In: Haddock A., & Wiseman R., (eds.), *The Anscombean Mind*. Routledge Philosophical Minds. London: Routledge, 241-252.

Rakoczy, H., Warnecken, F., and Tomasello, M. 2008. The Sources of Normativity: Young Children's Awareness of the Normative Structure of Games. *Developmental Psychology* 44 (3): 875–881.

Robinson, E., Haigh S., and Pendle J. 2008. Children's Working Understanding of the Knowledge Gained from Seeing and Feeling. *Developmental Science*, 11; 299-305.
Wright, C. 2012. Replies Part I: The Rule-Following Considerations and the Normativity of Meaning. In: Coliva, A. (ed.), *Mind, Meaning, and Knowledge: Themes from the Philosophy of Crispin Wright*. Oxford: Oxford University Press, 379–400.

Meyer, Marlene (University of Göttingen), Marina Proft (University of Göttingen), Lydia Paulin Schidelko (University of Goettingen), Jan M. Engelmann (University of California, Berkeley) and Hannes Rakoczy (University of Göttingen)

Social Metacognition – Preschoolers' Metacognitive Insight Under Partial Ignorance in Social Test Contexts

Being able to represent and explicitly reflect on their epistemic states (such as knowledge, ignorance, or uncertainty) marks a major milestone in young children's metacognitive development. Interestingly, these skills seem to develop relatively late. Prior research suggests that children are not able to reflect reliably on their ignorance until they are about 6 years old. Especially under so-called partial ignorance, when their knowledge is in fact limited, they struggle with acknowledging their uncertainty. Instead, they display a striking pattern of overconfidence as they overestimate their knowledge and incorrectly claim to know what is, e.g., hidden inside a box when they have previously been exposed to a set of possible objects (e.g., Robinson et al., 2006; Rohwer et al., 2012; Kloo et al., 2017). Typically, these experiments are conducted in individualistic contexts in which children need to state whether they do or do not know what the box contains.

However, this individualistic approach bears methodological as well as theoretical challenges. One considerable issue is that it neglects the pivotal link between metacognition and social interactions. The claim that metacognition has evolved for social purposes and plays a crucial role in social interaction and cooperation goes hand in hand with the claim that social interactions also enhance metacognition (e.g., Frith, 2012; Shea et al., 2014; Heyes et al., 2020). Therefore, a fundamental question arises: Do previous results underestimate children's true metacognitive competence because these experiments neglected the pivotal link between metacognition and its social functions?

We approach this question by introducing a social paradigm for testing children's ability to acknowledge their uncertainty under partial ignorance. In three studies, we asked whether 3- and 5-year-old children display successful metacognitive insight earlier in when given the opportunity to interact and communicate naturally with another person. In all three studies, children engaged with a second experimenter (E2) in a cooperative version of the partial ignorance task: E2 asked them for help in identifying which one of three toy animals has been hidden inside a box. In Study 1 (N = 66; 3- and 5-year-olds), E2 asked the children "Can you tell me which animal is inside the box?". Their answers were coded with regards to if and how they expressed uncertainty in the crucial test condition. In striking contrast to previous research, the majority of both 3- and 5-year-old children expressed uncertainty verbally and/or nonverbally, revealing that they were sensitive to their own ignorance under partial ignorance. Study 2 (N = 64; 3- and 5-year-olds) and Study 3 (in data collection, final N = 32; 3-year-olds) were conducted to increase commensurability with previous partial ignorance tasks and control for alternative explanations (e.g., Kloo et al., 2017; Rohwer et al., 2012). For instance, the crucial test question was phrased more explicitly ("Do you know which animal is inside the box?") and children's answers were coded more conservatively.

Preliminary results from these studies confirm the hypothesis that our social paradigm with its interactive and communicative design – tapping the collaborative nature of social interaction – facilitates children's ability to monitor, evaluate and explicitly assess their epistemic states: 3- and 5-year-old children communicated their uncertainty to E2 and acknowledged their ignorance in the crucial test conditions. The finding that even 3-year-olds were able to overcome their tendency to overestimate their knowledge presents a striking finding and will be discussed in light of the close relationship between metacognition and the communicative and cooperative character of human social interaction and cognition.

Michael, John (The University of Milan) and Stephen Butterfill (Warwick University)

The Moral Sense: A Minimal Approach

Moral psychology has seen a multi-decade attempt to adapt dual-process theories to the ethical domain. According to the leading dual-process theory, developed by Greene and Haidt (2002), fast processes (System 1) implement simple rules and characteristically arrive at deontological judgments backed by strong emotions, whereas slow processes (System 2) weigh the longer-term costs and benefits of available options to all affected parties and characteristically arrive at utilitarian judgments.

Tests of this theory has yielded conflicting results. For example, increasing time pressure has variously been found to make people's responses more deontological (Suter & Herwig, 2011), less deontological (Rosas & Aguilar-Pardo, 2020), and neither more nor less so (Bago & De Neys, 2019). The conflicting results provide an impetus to reconsider our approach to conceptualising fast processes in the ethical domain.

In particular, we suggest that it may be a mistake to think that fast processes involve any rules or norms at all. After all, given that fast processes have evolved to create benefits for survival and reproduction, we should not take for granted that they would have depended on the existence of norms or rules. Could there be moral mechanisms which operate quickly and without mediation by representations of norms or rules? If so, how would they operate and how would they have evolved?

To answer these questions, we draw upon a model of the evolution of mechanisms for detecting and avoiding toxins. These heterogenous mechanisms are observed in a great variety of animals, and presumably operate unmediated by representations of chemical laws. Applying this model to the moral domain, we must start out from the challenges which would have provided selection pressure for the evolution of moral mechanisms, and identify mechanisms which may have evolved to address these challenges without the need for representations of rules or norms. In the current paper, we illustrate this approach in relation to one key challenge, namely the need to manage social conflict. To this end, we examine a range of candidate mechanisms in humans and other primates, including dominance hierarchies and the emotions which promote reconciliation behaviours, both of which regulate conflict (Flack & De Waal, 2000). These cognitive, motivational and social mechanisms can serve to manage conflict without individuals having an awareness that this is their function or espousing this as a goal.

Crucially, the approach developed here provides the key to interpreting the aforementioned patterns of empirical findings: though fast moral mechanisms may influence participants' judgments in various ways in different settings, there is no reason to expect them systematically to lead to deontological judgments, since they do not involve deontological rules—nor utilitarian ones, nor indeed any rules or norms at all.

Michel, Matthias (New York University) and Ned Block (New York University)

Blank Access

Block (1995) distinguishes between phenomenal consciousness and access consciousness. Phenomenal consciousness is subjective experience. Access consciousness is the direct availability of information for executive functions, such as rational decision making, or verbal report. A central question is whether the two dissociate. Just as important is the methodological issue of determining how we could know whether they dissociate.

We examine a series of cases of access consciousness without phenomenal consciousness—'blank access', providing evidence for a difference between the two. While our putative blank access cases may not definitively prove a difference in nature, they do at least suggest that it is possible to empirically assess it.

Most of the discussion on access and phenomenal consciousness focuses on phenomenal overflow—purported cases of phenomenal consciousness without access consciousness. Some hold that it is not possible to empirically dissociate access and phenomenal consciousness, since subjects cannot provide

reports about overflowing phenomenal content (e.g., Kouider et al. 2007). As such, they conclude that theories hypothesizing dissociations between access and phenomenal consciousness "are inherently unfalsifiable and beyond the scope of science" (Cohen & Dennett, 2011, p. 358).

We argue for the dissociation, or at least for its empirical testability, by providing cases of 'blank access'—cognitive access without phenomenal consciousness. In blank access, a (perceptual) representation is accessed by executive systems without a phenomenal experience of its representational content. A representation might be, for instance, encoded in working memory and available for rational decision making, without there being anything it's like for the subject to experience the relevant representational content.

This way of testing the distinction has attracted surprisingly little attention. Yet, identifying cases of blank access provides a more straightforward way of settling the debate than finding out whether phenomenal consciousness overflows cognitive access.

Our list of putative blank access cases includes aphantasia, lag-1 sparing in the attentional blink, expert mental rotation, the case of reverse Anton syndrome, as well as reports from motion-induced blindness experiments. To remain brief, we only discuss the first two here. Subjects with aphantasia report having no visual imagery experiences. Aphantasia seems to be a perceptual effect, instead of stemming from response bias. For instance, visual imagery primes rivalry dominance in binocular rivalry—a phenomenon in which perception oscillates between incompatible images presented to each eye. But this priming effect is absent in aphantasia (Keogh & Pearson, 2018; Pearson, 2014). Similarly, visual imagery can also influence pupil dilation. But aphantasics do not show this effect (Kay et al. 2022). Finally, the oblique effect—a relative decrease in visual performance for oblique stimuli, is also typically present in tasks that require mental visual imagery, but absent in aphantasia (Keogh et al. 2021).

This apparent lack of visual imagery could indicate either a total lack of mental imagery, or unconscious visual imagery (Nanay, 2021). We argue for the latter: aphantasics can encode and manipulate unconscious iconic representations in working memory. Indeed, aphantasics show no accuracy deficits in standard imagery tasks (Liu & Bartholomeo, Preprint), including visual working memory tasks (Jacobs et al. 2018). In particular, Pounder et al. (2022) found no difference in performance between aphantasics and other subjects in a mental rotation task.

Furthermore, the linear relation between response time and angular difference between the figures—often interpreted as indicating the manipulation of a mental icon—was preserved in aphantasia.

There is evidence that aphantasics do not solve visual imagery tasks by using spatio-motor strategies, since performance on imagery tasks does not decrease when those strategies are unavailable (for instance, for imagining faces, or colors) (Liu & Bartholomeo, Preprint). Additional evidence comes from brain recordings. Weber et al. (Preprint) found no difference in visual working memory performance between weak and strong imagers. More importantly, decoding of visual contents from visual cortex was equally successful for both weak and strong imagers, and the decoded information predicted performance equally well across groups. Finally, Bates et al. (Preprint) analyzed the contralateral delay activity (CDA)—an event-related potential associated with the precision of visual working memory representation, in a visual working memory task across participants with varying levels of visual imagery vividness.

They found no correlation between vividness and performance, or between vividness and CDA, concluding that "the subjective sensory experience of [mental imagery] is distinct from the visual precision and capacity of [visual working memory]" (p. 33).

If this analysis is correct, aphantasia is a case of blank access. Aphantasics can encode, maintain, and manipulate representations in visual working memory to guide rational decisions in working memory tasks. Yet, there is nothing it's like for them to experience the contents they maintain and manipulate in working memory.

Here is a second putative case of blank access, which involves a phenomenon called 'lag-1 sparing' in the 'attentional blink'. When a subject tries to identify two targets in quick succession, if they identify the first target (T1), they often miss the second target (T2) if it appears within 200-500 milliseconds after it (Figure 1A, B). However, subjects can identify T2 if it occurs immediately after T1—a phenomenon known as 'lag-1 sparing' (Figure 1B). The main explanation for Lag-1 sparing is that T1 and T2 are jointly encoded in working memory without the need to deploy attention to T1 and T2 separately.

Recent studies found that lag-1 sparing preserves accuracy but not subjective visibility (or confidence) (Jones et al. 2020; Recht et al. 2019) (Figure 1 C and D). Accuracy is just as high for lag-1 and lag-6 conditions (where no attentional blink occurs), but visibility is just as low for lag-1 and lag-2 (where the attentional blink occurs). This suggests that, when presented at lag-1, T2 is as access conscious as when presented at

lag-6, while being just as phenomenally unconscious as when presented at lag-2. In other words, lag-1 sparing is a case of blank access—access consciousness without phenomenal consciousness.

Figure 1. A. The attentional blink paradigm. B. Typical attentional blink results showing lag-1 sparing. C. Visibility is equal at lag-1 and lag-2, but accuracy differs significantly. D. Confidence is equally low between lag-1 and 2, but accuracy is much higher at lag-1 (Adapted from Jones et al. 2020; Recht et al. 2019).

In summary, aphantasia and the lag-1 sparing effect could constitute cases of blank access. At the very least, these cases indicate that one can attempt to empirically establish the distinction between access and phenomenal consciousness. Of note, cases of blank access are also relevant for theories of consciousness. While blank access is compatible with theories like local recurrence theory, or higher-order theories, it is incompatible with global workspace theory.

Miłkowski, Marcin (Institute of Philosophy and Sociology, Polish Academy of Sciences)

Do problems with theory exacerbate the replication crisis?

There are widespread concerns regarding replicability, generalizability, cumulative nature of research, and cohesiveness of understanding of psychology and related fields (Simmons, Nelson, & Simonsohn, 2011; Open Science Collaboration, 2015; Boekel et al., 2015; Baker, 2016; Hughes, 2018; Yarkoni, 2022; Nosek et al., 2022). Questionable research practices, fraud, inappropriate methodological choices, and poor interpretability of research due to miscommunication are all contributing factors. The proposed reforms in psychology mainly focus on methodological issues, including rigorous statistical methods, preregistration, increasing transparency, encouraging experimental replication, and removing publication bias against null results (e.g., Chambers, 2017; Munafò et al., 2017; Nosek et al., 2022).

However, the unclear status and function of theory in research are suggested to be at the root of the crisis, making it a theory crisis (Klein, 2014; Muthukrishna & Henrich, 2019; Smaldino, 2017; Young, 2016; Carsel, Demos, & Motyl, 2018; Hughes, 2018; Oberauer & Lewandowsky, 2019; Szollosi & Donkin, 2019; Irvine, 2021; Hensel, Miłkowski, & Nowakowski, 2022; Levenstein et al., 2023; but see Trafimow & Earp, 2016 for a dissenting opinion).

This talk argues that theory-related issues may impact replicability and reproducibility of experimental results. Even if replication efforts fail due to fraud or other questionable research practices, inappropriate methodological choices, exacerbated by publication bias, or even due to underspecified designs in publications, theory-related factors are among the most plausible candidates to check for. Thus, theory-related representations play a crucial role in scientific inquiry, and their validity must be examined in case of replication failures. Even if better statistical standards, preregistration, transparency, culture of replication, and encouragement of journals to publish negative results are imposed, replication rates may still remain low.

In my view, a theory is a kind of cognitive artifact that serves to "maintain, display, or operate upon information in order to serve a representational function and that affect[s] human cognitive performance" (Norman, 1991, p. 11). The content of theories cannot be fully accounted for in terms of data, whether experimental or observational. Instead, theories provide a perspective on the phenomena under investigation. Multiple cognitive artifacts can form a stack that guides research practices. These artifacts may serve different functions and be tailored to the specific needs of a given domain (Miłkowski, 2022).

It is a common complaint that psychology lacks suitable theories (Oberauer & Lewandowsky, 2019; Fiedler, 2017; Muthukrishna & Henrich, 2019). My point is that it lacks various kinds of theories, including descriptive accounts of phenomena, as well as predictive, explanatory, and normative theories, which should not be conflated.

Predictive and explanatory theories constrain hypotheses put forward by researchers. The predictive value of a hypothesis depends to a large extent on the a priori probability of the hypothesis' being true: if most hypotheses we test are false, then it is only to be expected that a large proportion of empirically confirmed hypotheses are also likely false. Since it is theory that should constrain the pool of seriously entertained hypotheses, the suggestion is that a theory crisis might be at the root of the replication and generalizability crisis (Hensel et al., 2022). The rational evaluation of theories may be biased if only a limited number of alternative theories are entertained (Almaatouq et al., 2022; Dellsén, 2020).

Theories may thus fail to appropriately constrain predictive or explanatory hypotheses (Fiedler, 2017; Button et al., 2013; Bird, 2021; Muthukrishna & Henrich, 2019). Bird (2021), for instance, insists that the failures of replication should be expected in many fields of inquiry. Simply, theories in psychology are insufficiently robust to have a high prior probability.

Replication may also be hindered by inappropriate theoretical descriptions of observed phenomena (Eronen & Bringmann, 2021; Levenstein et al., 2023) and by confusing mere descriptions of the phenomena with their explanations (Scheel, Tiokhin, Isager, & Lakens, 2020). Gigerenzer (1998) observed that psychologists tend to use surrogates for proper theories: one-word explanations, redescriptions, muddy dichotomies, and data fitting.

Unfortunately, neither insufficiently robust theories nor surrogates for theory can be fixed by imposing methodological reform. Moreover, there may be deep reservations against theory development in various branches of psychology. Some areas of psychological research seem to have too little theory, while others rely on theory too much. On the one hand, researchers in many fields of psychology engage almost exclusively in gathering experimental and observational data, fishing for effects, even if they do not constitute targets of explanation (Cummins, 2000; Fried, 2020; van Rooij & Baggio, 2021). This actually promotes the atheoretical approach, which has led to the publication of an experimental paper in social psychology that contradicted the established principles of physics (Bem, 2011). When the statistical significance of the experimental effect becomes more important than the prior probability of the theoretical hypotheses, there is a significant pressure to design experiments that make no sense.

However, the atheoretical approach may be seen as a response to an overreliance on theoretical frameworks. Many existing theories seem irrefutable because they can accommodate almost any empirical finding (Frankenhuis, Panchanathan, & Smaldino, 2022; Miłkowski & Litwin, 2022; Roberts & Pashler, 2000; Szollosi & Donkin, 2019). Theories are also rarely if ever rendered obsolete (Ferguson & Heene, 2012). This should come as no surprise since many defenders of a theory focus on its confirmation, ignoring the need to compare it to competitors (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986). Sometimes researchers seem to find it sufficient to rephrase experimental results in their preferred theoretical vocabulary, rather than providing suitable experimental evidence that excludes competitors.

Overall, this suggests that issues with reproducibility, replicability, generalizability, and systematicity of research will persist without robust theoretical tools. The methodological reform is insufficient, but it can contribute to enhancing theoretical rigor. Appropriately rigorous theories could generate more plausible experimental hypotheses, exclude experimental designs that test for supernatural precognition, and provide a greater understanding of the phenomena under investigation.

References

Almaatouq, A., Griffiths, T. L., Suchow, J. W., Whiting, M. E., Evans, J., & Watts, D. J. (2022). Beyond Playing 20 Questions with Nature: Integrative Experiment Design in the Social and Behavioral Sciences. *Behavioral and Brain Sciences*, 1–55. doi: 10.1017/S0140525X22002874

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News*, *533*(7604), 452. doi: 10.1038/533452a

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. doi: 10.1037/a0021524

Bird, A. (2021). Understanding the Replication Crisis as a Base Rate Fallacy. *The British Journal for the Philosophy of Science*, *72*(4), 965–993. doi: 10.1093/bjps/axy051

Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, *66*, 115–133. doi: 10.1016/j.cortex.2014.11.019

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365. doi: 10.1038/nrn3475

Carsel, T., Demos, A. P., & Motyl, M. (2018). Strong scientific theorizing is needed to improve replicability in psychological science. *Behavioral and Brain Sciences*, *41*, e123. doi: 10.1017/S0140525X1800078X

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton: Princeton University Press.

Cummins, R. (2000). "How does it work" versus "what are the laws?": Two conceptions of psychological explanation. In F. Keil & R. A. Wilson (Eds.), *Explanation and Cognition* (pp. 117–145). Cambridge, Mass.: MIT Press.

Dellsén, F. (2020). The epistemic impact of theorizing: Generation bias implies evaluation bias. *Philosophical Studies*, *177*(12), 3661–3678. doi: 10.1007/s11098-019-01387-w

Eronen, M. I., & Bringmann, L. F. (2021). The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science*, *16*(4), 779–788. doi: 10.1177/1745691620970586

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological Science*, *7*(6), 555–561. doi: 10.1177/1745691612459059
Fiedler, K. (2017). What Constitutes Strong Psychological Science? The (Neglected) Role of Diagnosticity and A Priori Theorizing: *Perspectives on Psychological Science*, *12*(1), 46–61. doi: 10.1177/1745691616654458
Frankenhuis, W. E., Panchanathan, K., & Smaldino, P. E. (2022). Strategic ambiguity in the social sciences. *Social Psychological Bulletin*. Retrieved from https://www.psycharchives.org/en/item/e5bb9192-80a4-4ae4-9cda-5d144008196e
Fried, E. I. (2020). Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. *Psychological Inquiry*, *31*(4), 271–288. doi: 10.1080/1047840X.2020.1853461
Gigerenzer, G. (1998). Surrogates for Theories. *Theory & Psychology*, *8*(2), 195–204. doi: 10.1177/0959354398082006
Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, *93*(2), 216–229. doi: 10.1037/0033-295X.93.2.216
Hensel, W. M., Miłkowski, M., & Nowakowski, P. (2022). Without more theory, psychology will be a headless rider. *Behavioral and Brain Sciences*, *45*, e20. doi: 10.1017/S0140525X21000212
Hughes, B. M. (2018). *Psychology in crisis*. London: Macmillan Education : Palgrave.
Irvine, E. (2021). The Role of Replication Studies in Theory Building. *Perspectives on Psychological Science*, *16*(4), 844–853. doi: 10.1177/1745691620970558
Klein, S. B. (2014). What can recent replication failures tell us about the theoretical commitments of psychology? *Theory & Psychology*, *24*(3), 326–338. doi: 10.1177/0959354314529616
Levenstein, D., Alvarez, V. A., Amarasingham, A., Azab, H., Chen, Z. S., Gerkin, R. C., … Redish, A. D. (2023). On the Role of Theory and Modeling in Neuroscience. *Journal of Neuroscience*, *43*(7), 1074–1088. doi: 10.1523/JNEUROSCI.1179-22.2022
Miłkowski, M. (2022). Cognitive Artifacts and Their Virtues in Scientific Practice. *Studies in Logic, Grammar and Rhetoric*, *67*(3), 219–246. doi: 10.2478/slgr-2022-0012
Miłkowski, M., & Litwin, P. (2022). Testable or bust: Theoretical lessons for predictive processing. *Synthese*, *200*(6), 462. doi: 10.1007/s11229-022-03891-9
Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. doi: 10.1038/s41562-016-0021
Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*, 221–229. doi: 10.1038/s41562-018-0522-1
Norman, D. A. (1991). Cognitive Artifacts. In J. M. Carroll (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface* (pp. 17–38). Cambridge: Cambridge University Press.
Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., … Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*(1), 719–748. doi: 10.1146/annurev-psych-020821-114157
Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, *26*(5), 1596–1618. doi: 10.3758/s13423-019-01645-2
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). doi: 10.1126/science.aac4716
Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–358.
Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2020). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science*, *16*(4), 744–755. doi: 10.1177/1745691620966795
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. doi: 10.1177/0956797611417632
Smaldino, P. E. (2017). Models Are Stupid, and We Need More of Them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational Social Psychology* (1st ed., pp. 311–331). New York: Routledge. doi: 10.4324/9781315173726-14
Szollosi, A., & Donkin, C. (2019). Neglected Sources of Flexibility in Psychological Theories: From Replicability to Good Explanations. *Computational Brain & Behavior*, *2*(3–4), 190–192. doi: 10.1007/s42113-019-00045-y
Trafimow, D., & Earp, B. D. (2016). Badly specified theories are not responsible for the replication crisis in social psychology: Comment on Klein. *Theory & Psychology*, *26*(4), 540–548. doi: 10.1177/0959354316637136
van Rooij, I., & Baggio, G. (2021). Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 682–697. doi: 10.1177/1745691620970604
Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, e1. doi: 10.1017/S0140525X20001685
Young, G. (2016). *Unifying Causality and Psychology*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-24094-7

Mölder, Bruno (University of Tartu)

An interpretivist perspective on emotions

The aim of this talk is to delineate the implications of interpretivism for the study of emotions. Intepretivism is a general approach to the mind on which the possession of mental properties is constitutively dependent on a certain kind of ascription. It has not been explicitly developed for emotions, and it should be distinguished from psychological constructionism about emotions, including its interpretivist variants. An important aspect of interpretivism is that it takes the mind and mental states to be "shallow", that is, circumscribed by our folk psychological conception. It follows from this assumption that mental kinds do not have a deeper essence that would lay beneath the folk psychological classification. Different versions of interpretivism may approach the ontology of the mental differently, but in my preferred version, mental properties exist only in a deflationary, pleonastic sense. In this sense, talking about the properties of something is just another way of talking about what predicates can be applied to it. Given this, the individuation of mental kinds is not independent of us. It depends on our interests and purposes in social interaction and on the particular outlook of our folk psychology. The same applies to emotions insofar as they are taken as mental kinds.

I argue against the idea that emotions are natural kinds, as far as emotions are understood through folk psychology. Natural kinds are classes of homogeneous entities that are individuatively-independent from our minds and that support inductive generalizations. I discuss some reasons for rejecting the natural kind status of emotions. That includes the cultural variation of classifications, the regulative role of folk psychology, holism and vague boundaries between emotions. Conceiving emotions through folk psychology and not as natural kinds has implications for the study of emotions, for it makes sense to focus more on the (conceptual and empirical) study of our folk conceptions.

That leaves open the issue of the relationship between folk emotion terms and the related theoretical notions used in affective science (such as "core affect", affect program", "emotional episode", etc.). On the latter, interpretivism does not preclude empirical work. However, do folk and scientific notions refer to the same thing? I discuss possible positions an interpretivist could take in this regard: global interpretivism, restricted interpretivism, and extending folk psychology. I conclude that the kinship of folk notions of emotion and scientific notions is more a matter of policy than a fact. It depends on whether we are interested in emphasizing the continuity between the folk conception and its scientific refinements.

Molina, Alonso (University of California, Los Angeles)

Spontaneous and Voluntary Introspection

Many philosophers and scientists (see Müller, 1911; Armstrong, 1963, 1968, 1999; Lycan, 1995; Hill, 1988; Sosa, 1998; Prinz, 2004; Renero, 2019) have distinguished between two kinds of introspection. The first kind is typically described as automatic and fast, while the second kind is described as deliberate and effortful. Think about the difference between an agent stubbing her toe and shouting, "Oh! I am in such pain," in contrast to her assessment of her ongoing abdominal pain such as, "I am suffering from a radiating sharp pain at a 7 on this pain scale," when asked by ER doctors. In the former case, there is no previous deliberation to reach the introspective judgment "Oh! I am in such pain." In the latter case, on the other hand, her introspective assessment requires deliberate and focused attention on the pain in order to, hopefully, provide an accurate description of it. Similarly, on many occasions, we are suddenly seized by emotions and say to ourselves, for example, "I am so happy right now," while other times we introspect the causes of our emotions or focus on the phenomenological character of our emotional states, e.g., think of a writer describing what it is like to feel fear.

Now, despite the broad acknowledgement of two kinds of introspection by many philosophers, no one has given a clear answer as to how we should account for these two kinds. Instead, they have chosen

to prioritize one kind of introspection over the other in their accounts. The most notorious examples are Armstrong's (1963. 1968, 1999) and Lycan's (1995) accounts of introspection that put the fast, automatic kind at the center of their theory while neglecting the deliberate, effortful kind of introspection. On the other hand, many philosophers do not acknowledge the automatic kind of introspection, and, instead, assume that all introspection is necessarily effortful and deliberate. Rosenthal (1986), for example, claims that introspection requires a deliberate action of focusing on a conscious mental state. There are also philosophers who seem to be agnostic on the matter. For example, Gertler's (2000; 2001) acquaintance view of introspection, and Goldman's (2005) multi-monitoring account of introspection do not explicitly specify if they take introspection to be deliberate or not.

In what follows, I aim to provide an account of the two kinds of introspection in light of their attentional mechanisms. The automatic kind of introspection will be called spontaneous introspection, while the deliberate, effortful kind will be called voluntary introspection. I argue that we have both of these kinds of introspection – I call this the dual account of introspection. I distinguish between the two kinds of introspection on the basis of the predominant attentional process behind each. My contribution lies in mapping two subcategories of internal attention (see Chun et al., 2011) with both the automatic and the deliberate kind of introspection identified by philosophers. These subcategories are internal bottom-up attention and internal top-down attention (see Legrain et al., 2009), respectively.

Although attention is understood to play a crucial role in introspection by many accounts (see Armstrong, 1963, 1968, 1999; Lycan, 1995; Hill, 1988; Gertler, 2000,20001; Rosenthal, 1986, 2000), the introspective literature has not always been explicit regarding what type of attention is involved in introspection. It is acknowledged that attention allows agents to select, prioritize, and maintain focus on their ongoing mental states, i.e, the introspected states, thus giving rise to an introspective state. However, it is not clear what categories or types of attention are relevant to introspection nor how the relation between attention and introspection works.

Thus, in order to account for the two kinds of introspection, I first argue that internal attention (also called intellectual attention) is the relevant type of attention at work in introspection. Doing so will require me to explain what internal attention is, distinguish it clearly from perceptual/external attention, and explicate its role in introspection. After doing this, I argue that two subcategories of internal attention, namely, internal bottom-up attention and internal top-down attention, account for spontaneous and voluntary introspection, respectively.

Bottom-up and top-down attention are divisions of attention concerned with how much control an agent can exert over their attention. Typically, bottom-up attention is considered to be exclusively externally-oriented. However, in my argument, I explain that only by incorporating both bottom-up and top-down subcategories into internal attention can phenomena such as spontaneous introspection, spontaneous remembering, spontaneous mind-wandering be explained. I support my claim with empirical evidence (cf. Amir et al, 2021; Bernsten, Vyshedskiy, 2020; Irving, 2016; among others). Framing introspection in terms of its attentional mechanisms will provide us with the theoretical tools to explain cases of both spontaneous and voluntary introspection.

I think that dissecting introspection in this way will help elucidate its nature, its relation with attention, and answer some problems currently debated in the philosophical literature. Particularly, it will help us answer if introspection is effortful, explain why –on some occasions– agents seem to introspect spontaneously, and provide us with significant insight concerning the relationship between self-knowledge and introspection.

The outline of my paper is as follows: in the first part of the paper, I motivate my account. I start by defining some important terms and providing a general characterization of introspection. The second part of the paper is devoted to understanding attention's role in introspection. Here I discuss internal and external attention. After arguing for internal attention as the main attentional mechanism at work in introspection, I argue that two of its subcategories, namely, internal bottom-up attention and internal top-down attention, account for spontaneous and voluntary introspection, respectively. In the final part of the paper, I provide my full account of introspection and present relevant cases. After that, I discuss the advantages of my account over other accounts of introspection and discuss related issues such as the connection between introspection and self-knowledge in light of my dual account.

Myers, Joshua (LOGOS University of Barcelona)

The Epistemic Role of Vividness

1. Introduction

Mental images come in different degrees of vividness. Some images are highly forceful, rich, and 'lifelike,' while others are faint, blurry, and impoverished. But although vividness has been a topic of discussion in philosophy for centuries, very little attention has been paid to the epistemology of vividness.

This is surprising, since it is very plausible that vividness plays an epistemic role. Suppose you are wondering whether your luggage will fit in the trunk of your car. Intuitively, a vivid, clear, and precise mental image of your luggage fitting comfortably into the trunk gives you more justification for believing that the luggage will fit than a faint, blurry, and sparsely detailed imagining.

So, vividness is epistemically relevant. But how? I will argue that vividness is higher-order evidence about one's own epistemic state, rather than first-order evidence about the world. First, I will argue against the first-order view. Next, I will argue for the higher-order view.

2. Against Vividness as First-Order Evidence

We can formulate the thesis that vividness is first-order evidence as follows:

Vividness as First-Order Evidence (VFO): Mental images with a greater degree of vividness can confer a greater degree of justification.

I will pose a dilemma for VFO: either vividness is a matter of content, or it is a matter of non-contentful phenomenology. Neither way of cashing out the notion of vividness ultimately vindicates VFO.

On the first horn, vividness is a matter of content, or what is represented by the imagining. Let us call this conception of vividness content-vividness.

Plausibly, for an image to (non-inferentially) justify a belief that p, it must itself represent that p. Thus, everyone can agree that how content-vivid an image is will determine which beliefs it can justify. But VFO says that vividness determines not just which beliefs are justified but the degree to which they are justified. It is much less plausible that content-vividness can play this role. As long as your imagining is content-vivid enough to represent your luggage fitting into the trunk, then no amount of extra content-vivacity will increase the amount of justification you get for this belief. Adding in fine-grained detail with respect to the color, texture, and shape of the suitcase might allow the imagining to justify additional beliefs about the suitcase, but it will not give you any additional justification for the belief that the luggage will fit in the trunk.

On the second horn of the dilemma, vividness is a matter of non-representational phenomenology, or how an imagining feels independently of what it represents. Let us call this conception of vividness phenomenal-vividness.

It is difficult to see why the phenomenal-vividness of an image gives one any more reason to believe that it is accurate. Many inaccurate imaginings are highly phenomenally vivid. Conversely, subjects that score low on measures of vividness are able to perform just as quickly and accurately as normal subjects on reasoning tasks that implicate imagery, such as mental rotation, mental scanning, and visual recognition (Berger & Gaunitz 1977, Dean & Morris 2003, Pounder et al. 2022). Instead, phenomenal-vividness correlates with factors that are epistemically irrelevant. For example, vividness of mental imagery correlates with background music (Martarelli et al. 2016, Herff et al. 2021) and emotional valence (Bywaters, Andrade, & Turpin 2004, Bohanek, Fivush, & Walker 2005). But neither the presence of background music nor the emotional valence of an image's content should impact the justificatory force of a mental image.

So, on any plausible way of precisifying the notion of vividness, VFO is false.

3. Vividness as Higher-Order Evidence

Instead, we should account for the intuition that vividness is epistemically relevant by affording it a role as higher-order evidence.

There is a robust correlation between the vividness of a mental image and the prior information one has about its subject matter. Multiple studies find that events are imagined more vividly when they take place in familiar locations (Arnold, McDermott, & Szpunar 2011, Robin & Moscovitch 2014). Other studies have extended this finding and found that vividness corresponds to the familiarity of content elements beyond

just location, such as people and objects (Baddeley & Andrade 2000, Argembeau & Van der Linden 2012). This coheres well with empirical models of vivacity that understand it as an index of how much sensory information is available in long-term memory. Baddeley & Andrade 2000 found that disrupting long term memory decreases vividness of imagery. More recently, D'Angiulli et al. 2013 found that vividness of imagery predicts success at incidental recall. The authors conclude that "vividness may act as an index of availability of long-term sensory traces" (p. 1).

The correlation between vividness and prior information allows vividness to play a role as higher-order evidence. The vividness of an image is evidence about the amount of prior evidence one has about its subject matter. If you find that you can form a highly intense, detailed, and precise imagining of the suitcase fitting in to the trunk, then this is some evidence that you have a large pool of information regarding the suitcase and trunk to draw on in constructing that imagining. By contrast, if you find that are only able to form a faint, hazy, and imprecise imagining of the suitcase fitting into the trunk, then this indicates that you do not have much evidence regarding the suitcase and trunk.

However, there is an asymmetry in the kind of higher-order evidence that vividness can provide. This is because it only acts as a proxy for the amount of prior information one has, rather than the epistemic status of that information. For example, you might have a rich set of prior beliefs that results in a highly vivid image, but those beliefs might be unjustified and therefore incapable of conferring justification. So, a high level of vividness is not evidence that an image is based on adequate evidence, even though suitably low vividness is evidence that an image is not based on adequate evidence.

Niczyporuk, Aneta (University of Bialystok) and Edward Nęcka (Jagiellonian University)

The relationship between thought suppression research paradigms

Research on thought suppression, understood as volitional blocking selected mental content from entering the awareness, is centered around two paradigms: the white bear (WB) paradigm and the think/no-think (TNT) paradigm. Depending on which of these two procedures one conducts research in, different conclusions about the effectiveness of thought suppression will emerge. While research in the WB paradigm points to the paradoxical effects of thought suppression, TNT studies indicate that it is effective.

There are many differences between the two thought suppression tasks. In the WB procedure, participants are asked to suppress a single content. At the same time, they are supposed to signal if a forbidden thought pops into their consciousness. Typically, after suppression, participants experience more target thoughts than participants in the control group who did not suppress. In TNT, on the other hand, participants initially learn dozens of words. In the next stage, they see cues of some of these words and are asked either to recall or to suppress a thought of a particular word. At the very end, the memory of all the words is tested, and it is usually found that the suppressed words are recalled more poorly than the control words.

The differences in the methodology of the thought suppression tasks are considerable, but if the two procedures study the same phenomenon, their outcomes should be consistent. However, it is possible that the tasks only ostensibly are about the same thing, while the actual processes involved in dealing with unwanted thoughts in each procedure are distinct. In our study, we tested the hypothesis that thought suppression in the WB task is carried out by the resistance to proactive interference (cf. Bomyea & Amir, 2011; Ólafsson et al., 2013; Wessel et al., 2008), while suppression in the TNT task is carried out by prepotent response inhibition (cf. Anderson & Green, 2001; Anderson & Huddleston, 2012; Visser et al. 2020). According to this hypothesis, the WB task would represent situations where unwanted thoughts arise for no apparent reason, and the TNT task would relate to circumstances where there is an external stimulus provoking intrusions. In addition to testing the presented hypothesis, our study also aimed to investigate the relationship between the two thought suppression tasks. Concurrently, we also wanted to replicate Friedman and Miyake's (2004) inhibitory function model.

In our research, participants went through both thought suppression procedures. In the following study session, they performed tasks on two executive functions: prepotent response inhibition (Stroop task, stop-signal task, antisaccade task) and resistance to proactive interference (Brown-Peterson task, AB-AC task, cued-recall task).

Thought suppression procedures appeared to be related, but only in the case of intrusion frequency during thought suppression phases. This may indicate that in both procedures during suppression attempts there are engaged some common processes. Friedman and Miyake's (2004) inhibitory function model was replicated with one difference: Prepotent response inhibition and resistance to proactive interference were strongly related, which may suggest that they should be considered as a single function. The results did not support the hypothesis that each of the thought suppression procedures uses different inhibitory functions. However, the lack of correlation between tasks on thought suppression and executive functions does not mean that these tasks are performed by entirely different processes (cf. Apšvalka et al., 2022). In the presentation, possible explanations will be discussed and the results will be related to previous studies. Methodological problems of research on processes involved in mental control in different contexts will also be considered.

Literature:

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control.
Nature, 410(6826), 366-369.

Anderson, M. C., & Huddleston, E. (2012). Towards a cognitive neurobiological model of motivated forgetting. In R. F. Belli (Ed.), True and false recovered memories: Toward a reconciliation of the debate (pp. 53–120). Springer Science + Business Media.

Apšvalka, D., Ferreira, C. S., Schmitz, T. W., Rowe, J. B., & Anderson, M. C. (2022). Dynamic targeting enables domain-general inhibitory control over action and thought by the prefrontal cortex. *Nature communications*, *13*(1), 274.

Bomyea, J., & Amir, N. (2011). The effect of an executive functioning training program on working memory capacity and intrusive thoughts. Cognitive Therapy and Research, 35, 529-535.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. Journal of Experimental Psychology: General, 133(1), 101-135.

Ólafsson, R. P., Emmelkamp, P. M., Gunnarsdóttir, E. R., Snæbjörnsson, T., Ólason, D. T., & Kristjánsson, Á. (2013). Suppressing disgust related thoughts and performance on a subsequent behavioural avoidance task: Implications for OCD. Behaviour Research and Therapy, 51(3), 152-160.

Wessel, I., Overwijk, S., Verwoerd, J., & de Vrieze, N. (2008). Pre-stressor cognitive control is related to intrusive cognition of a stressful film. Behaviour Research and Therapy, 46(4), 496-513.

Visser, R. M., Anderson, M. C., Aron, A., Banich, M. T., Brady, K. T., Huys, Q. J. M., Monfils, M. H., Schiller, D., Schlagenhauf, F., Schooler, J. W., & Robbins, T. W. (2020). Neuropsychological mechanisms of intrusive thinking. In P.W. Kalivas & M.P. Paulus (Eds.), *Strüngmann Forum Reports* (pp. 124-184). The MIT Press.

Niemeck, Maik (Philipps-Universität Marburg)

Indexical Thought, Affordances and Virtual Worlds

In this talk, I will explore the idea that indexical thought typically comes with an awareness of affordances relations that hold between the thinker and the entities thought of (including himself) (Prosser 2015). Indexical thoughts represent objects, places and times within actionable reach. When we, for instance, think first-personally about an object it is given to us as the thing through which we can perform privileged actions, i.e. actions most other objects apparently cannot perform through it and we cannot perform through other objects, such as moving the arm purely by will. When I grasp thoughts about things other than myself, then I am aware of them as entities I can merely interact with, but not as things I can do something through, i.e. as something over which I have immediate control. Similarly, if we think of a place as "here" rather than "there", we think of it as a place where we can do stuff, since thinkers are usually at the place of their thoughts. Thinking of a time as "now" will in a similar fashion involve an awareness of the possibilities this time provides to the thinker – that thinking of times as "tomorrow" or "yesterday" presumably lacks. In respect thereof, indexical thoughts usually entail certain affordances presuppositions.

I will show that this idea is also supported by our interaction with virtual worlds. When playing video games or using virtual reality headsets, people tend to speak about the worlds presented by these devices in indexical terms, suggesting that they are grasping indexical thoughts about them. They say things like "I died

in the boss fight" or "I am here at the lake" while being fully aware that their usual biological body is still intact and presumably sitting in an armchair. This is very different when people are passively watching TV shows or reading a book and talking about both. I will conclude that the crucial difference between both ways of thinking about these virtual worlds is grounded in the ways they provide action possibilities to an agent and that all this is strong evidence for the claim that indexical thought is closely connected to the representation of action possibilities – a view that has been challenged by so-called de se skeptics (Cappelen & Dever 2013; Magidor 2015). Hence, the talk aims to defend three main claims: A) indexical thoughts typically come with an awareness of affordances relations; B) claim A is supported by our interaction with different types of virtual worlds; and C) claim A can also be used to refute the position known as de se skepticism.

The first part of the presentation is concerned with some theoretical preliminaries and the scope of claim A. To this end, I will introduce some necessary distinctions – such as the difference between pure indexicals and true demonstratives (Kaplan 1989), that between virtual worlds and virtual realities (Chalmers 2022), and that between affordances and the perception of affordances (Gibson 1977; Nanay 2011, Siegel 2014) – and discuss some general problems that my approach prima facie has to cope with. For instance, it seems obvious that not all indexical thoughts typically involve an awareness of affordances. In fact, some indexical thoughts might even come with an awareness of the absence of possibilities to interact with the objects of thought. When I think that something is over there or something happened yesterday, these states of affairs are in most, perhaps all, circumstances presented to me as something I have no chance of changing. So, it is a complicated endeavor to single out a class of indexical thoughts which share affordances presuppositions and to find out what unites them. I will concentrate on the most paradigmatic types of thought that we entertain about our current situation, namely thoughts expressed using (non-anaphorically, non-demonstratively, not deferred etc.) the pure indexicals "I", "here" and "now".

The second section of the talk explains what the central questions in the debate on de se skepticism are and why many philosophers (Babb 2016; Bermudèz 2017, 2018; García-Carpintero 2016; Niemeck 2022, Owens 2011; Torre 2016, 2017; Valente 2017) have claimed that (certain types of) actions require indexical thoughts. In contrast to these accounts, I will propose that we don't have to assume such a strong modal relation between indexical thought and action to address the challenges posed by the skeptics and that we can instead refute de se skepticism by exploring neglected features of indexical thinking. One of these features is that some indexical thoughts typically include an awareness of affordances that non-indexical thoughts usually lack. This makes indexical thought stand out and might explain why philosophers thought that action necessitates indexical thought. All of this becomes clear when observing indexical thinking in the wild.

The final part seeks to do this and single out the features of virtual worlds that cause us to think indexically about them. I will briefly discuss questions such as of what kind of affordances awareness (e.g. doxastic or perceptual) is most likely to be involved in indexical thinking, what sort of relation holds between an awareness of affordances and indexical thinking (e.g. grounding or causing), and what the relevant features of virtual worlds (e.g. immersion or interactivity) could be that cause these thoughts about them. For this purpose, I will compare media that present highly immersive but little interactive worlds (e.g. VR-movies) with those that present highly interactive but little immersive worlds (e.g. early text adventure video games). I will demonstrate that only interactive virtual worlds seem to trigger indexical thoughts and that immersion matters only insofar as it creates a feeling of being in touch with the world. The latter also suggests that an awareness (understood in a nonfactive way) of affordances is crucial for indexical thinking and not affordances themselves. Finally, I will also address the question of whether indexical thoughts trigger an awareness of affordances or if indexical thoughts are triggered by an awareness of affordances.

Nouwen, Rick (Institute for Language Sciences, Utrecht University)

Reconciling the philosophy and cognitive science of rhetorical figures: the case of meiosis and hyperbole

One major development in the scientific study of linguistic meaning in the past 15 years is the use of cognitive models of pragmatic behaviour (Goodman and Frank, 2016). Such models are connected to

philosophy in two ways. First of all, they find their origins in game-theoretical approaches to cooperative behaviour (Lewis, 1969; Franke, 2009). Second, their applications concern phenomena that have been central to discussions in philosophy of language. Here, I turn to the pragmatics of rhetorical figures and explore the relation between philosophical approaches and approaches from the cognitive modeling tradition.

The figures of speech I am concerned with are meiosis and hyperbole. These are deliberate, yet cooperative over- and under-statements that are usually (but not always) false. Consider the following scenario as illustration. Timid gave a housewarming party last week. He invited 60 people, expecting only few to come. Astonishingly, all 60 people turned up. Scarlett, who was at Timid's party, meets up with Timid afterwards and is surprised to find Timid rather insecure about whether his party was a success. Consider the following two sentences, Scarlett could utter in response.

(1) There were 100 people in your living room!

(2) Your living room was completely empty!

By itself these may seem infelicitous in this situation, but with some contextual clues they become useful rhetorical moves for Scarlett. The example (1) is a case of hyperbole, Scarlett exaggerates the number of party goers in order to make the point that Timid should be happy: "Are you kidding me? Of course it was a success! There were 100 people in your living room!". For (2), a case of meiosis, you should imagine Scarlett adopting a mocking sarcastic tone: "You are right. The party was a total disaster. Your living room was completely empty!".

Interestingly, if we change the context, the rhetorical role of the sentences in (1) and (2) changes as well. Imagine Brag, who also threw a party, also inviting 60 people and expecting half to come. In reality, only 20 people came. Yet, when Brag talks to Scarlett (who attended Brag's party) he confidently boasts what a success the party was. Now, Scarlett can counter Brag's boasting using the examples above. This time, the sarcastic tone goes with (1) and is absent from (2). Walton (2017) and Author (2023) argue that in Brag's context (1) is a case of meiosis, while (2) counts as hyperbole. They define hyperbole as exaggerating an existing deviation between a norm and reality, while meiosis involves presenting this deviation as non-existent or going in the opposite direction. In this way,these theories argue that meiosis and hyperbole are essentially scalar phenomena, where propositional contents are ordered and we have intuitions about distances between such contents. Empirically, these theories are grounded in the fact that meiosis seems to involve a specific kind of verbal irony, often called impersonation or pretence irony (see, Currie 2006; Simonin 2018).

Philosophical approaches to rhetorical figures like the ones just discussed have as one of their main goals to understand the nature of tropes. In particular for meiosis/hyperbole, they aim to understand how the propositional content of false statement can still be communicative and, more specifically, what is the nature of verbal irony (e.g. Grice 1978; Sperber and Wilson 1981; Wilson 2006). This is contrasted by recent influential studies presenting cognitive models of rhetorical figures (Kao et al., 2014; Yoon et al., 2020), where the focus is on accurately predicting (experimental) data on how interlocutors balance the utility of truthful versus untruthful use of language. Phenomena like irony, hyperbole and meiosis are modeled using Bayesian probabilistic models in the Rational Speech Act framework (Goodman and Frank, 2016; Scontras et al., 2021). For instance, Kao et al. (2014) propose that hearers do not just reason about the speaker's beliefs about the world, but they also reason about her communicative goals. In particular, interpreting an utterance involves reasoning about what kind of question under discussion the utterance was intending to resolve. On Kao et al.'s theory, sentences with a meaning that is a priori highly unlikely heighten the posterior probability that the speaker does not (just) intend to provide information about the world, but (also) intends to convey something else, e.g. some affective meaning.

While models Like Kao et al.'s manage to predict non-literal interpretations, they fail to derive inferences that hearers could draw from figures like hyperbole. Figure I illustrates a simulation using Kao et al's model of an overstatement. The model correctly predicts a high probability for a non-literal interpretation, but only derives that this interpretation is likely to be somewhat removed from expectations – that it is somewhat unusual.

Note that the cognitive models are, in fact, very close to the philosophical approaches discussed above in that they assume that interlocutors reason about the propositional content of an utterance even if this content is highly likely to be false. Also, they compare this content to other contents on a scale and compare this to norms (Bayesian priors) on that scale. What is missing from these models is exactly something that the philosophical approaches can bring, namely a notion of truthlikeness. Given the scalar setup, not all false statements are alike. Some untruths are further from the truth than others. In a cognitive model this can influence the utility of a false message. That is, deviation from the truth is costly, but bigger deviations are more costly than smaller ones. Using this intuition, we can improve probabilistic models. Factoring in

this cost, Kao et al's model correctly derives the hyperbole inference that the truth is somewhere between the norm and the literal meaning. See figure II.

In summary, I propose an fruitful exchange of ideas between philosophical and modeling approaches to the rhetorical figures of meiosis and irony.

Openshaw, James (Université Grenoble Alpes & Ruhr-University Bochum)

(In defence of) preservationism and the previous awareness condition: What is a theory of remembering, anyway?

I argue that the theories of remembering one finds in the recent philosophical literature—simulationist/functionalist, causalist, and epistemic—are best characterised as answers to questions posed at three distinct levels of inquiry.

(Q1) Under what conditions does remembering occur? (The psychofunctional question.)

(Q2) Under what conditions is there some event in one's personal past e such that one is remembering e? (The reference question.)

(Q3) Under what conditions is there some event in one's personal past e such that one is accurately remembering e? (The accuracy question.)

Simulationist (Michaelian 2016) and functionalist (Fernández 2019) views, whose focus is (Q1), are best seen as theories of psychofunctional process types. Causalist views (Martin & Deutscher 1966), whose focus is (Q2), are best seen as theories of referential remembering. Epistemic views (Hoerl 2022), whose focus is (Q3), are best seen as theories of successful remembering. Insofar as there is conflict between these theories, it is a conflict of integration rather than—as widely presented—head-on disagreement. In short, the leading theories of 'what remembering is' have, as their principal focuses, overlapping but crucially distinct subject matters.

This view of the landscape has two benefits. First, though it does not dissolve disputes about the nature of remembering by casting them as purely verbal, it clarifies the dialectical rules of engagement and illuminates a path to integration. Once these different projects are demarcated and held in view, we can move past seeing the theories as in direct competition. We can then focus on identifying conflicts of integration, moving the debate forward by pursuing a harmony across these various levels of theorising.

Second, we can see two influential principles, the previous awareness condition and preservationism, as principles concerning reference and accuracy in remembering, respectively.

Previous awareness condition (PAC): One can remember some particular—an event, object, sensation, etc.—only if there is, in one's personal past, an experience in which one was (non-mnemically) aware of it.

Preservationism: One can remember some particular as having been F only if there is, in one's personal past, an experience in which one was (non-mnemically) aware of its being F (or some more determinate way F').

Where either has been rejected (De Brigard 2014; Michaelian 2022), it is, I argue, due to arguments which slip between these different levels of theorising. PAC articulates a necessary condition for referential remembering. Preservationism articulates a necessary condition for accurate remembering. These principles are perfectly compatible with constructivist, anti-transmissionist theories of the psychological mechanisms underpinning episodic remembering.

Otterski, Emma (University of Edinburgh)

Do we directly perceive others' emotions?

Having a grasp of others' emotions can be important for acting and interacting appropriately. Sometimes it is necessary for survival. Despite this, emotion attribution, or 'recognition' as per the psychological literature,

has received relatively little attention in the philosophical mindreading literature, though it is generally thought to fit into either a theory-based or simulation account. For some authors, however, emotions reveal that standard accounts of mindreading are misguided. The focus on belief, and particularly false belief, in the mindreading literature – a non-veridical, epistemic state – may have led to underestimating the role of perception in attributing mental states such as emotions, which often involve visible and easily recognisable facial expressions. On seeing someone crying we seem to see that they are sad, on seeing someone laughing we may think that we see their happiness. Given these insights, direct social perception (DSP) theories (e.g., Gallagher, 2008; Gallagher and Varga, 2014; Zahavi, 2011; Newen, Welpinghus and Juckel, 2015) propose that emotions can be directly perceived as opposed to inferred.

A well-documented problem for evaluating DSP accounts is that it is not always clear what is meant by non-inferential perception. Drawing on Drayson's (2018) work on the inferential/non-inferential distinction in the wider direct perception literature, I first suggest that proponents of DSP are sometimes talking about psychological directness, sometimes about epistemological directness, and sometimes metaphysical directness. This is interesting in itself and helps to clear up some of the ambiguities in the DSP/mindreading debate. While initially put forward as an alternative to standard mindreading accounts, and so most naturally read as denying intra-perceptual inferences (and so advocating psychological directness), current DSP proponents tend to offer a more conciliatory approach, in which they spell out the inferential/non-inferential distinction differently while still maintaining that this sheds some light on the psychological processes underlying our ability to grasp others' emotions.

A promising way to spell this out is through comparison to object recognition; seeing the person as happy is like, and no more inferentially complex than, seeing the mug on the table. Neufeld (2020) and Newen et al. (2015) each draw on the similarity of emotion recognition and objection recognition to offer an account of DSP. Using research in visual categorisation, Neufeld argues that emotion concepts are the basic-level concepts applied to facial expressions in perception, and it is in this sense that emotions are perceived similarly to ordinary objects and 'non-inferentially'. Newen et al. argue that emotion recognition is a form of pattern recognition that is functionally analogous to object recognition – on their account, it looks like the psychological processes are suitably non-inferential in a way not recognised by mindreading accounts, and there is also non-inferential epistemic access to others' emotions.

Against Neufeld, I argue that categorisation of affective facial expressions is sometimes significantly different to ordinary object categorisation. Drawing on work in social psychology, I suggest that emotion categorisation is more flexible than ordinary object categorisation, meaning the entry-level (first) concept applied is more variable than for ordinary objects. While these insights do not undermine the idea that basic emotion concepts are sometimes the entry-level concepts applied in visual categorisation, it does throw doubt on whether they play the same role as basic-level concepts do for object categorisation in real-world contexts and how similar recognising affective expressions is to recognising objects. While Newen et al.'s pattern recognition account can accommodate some contextual factors that Neufeld's cannot, I suggest that it still doesn't take into account the full flexibility of emotion perception because of its reliance on recognition of the so-called basic emotions. On top of this, the process they outline looks sufficiently different to ordinary object recognition due to the spatiotemporally extended nature of the emotion pattern's token parts. The comparison to ordinary object recognition, then, does not account for the differing effect of context and goals on social concept application as opposed to the concepts applied to ordinary objects. Facial expression perception is flexible; affective expressions are sometimes perceived as emotions, but sometimes the same expression is perceived not as an emotion but still as an expression with affective significance. This opens the door to distinguishing between emotion perception/recognition and affect perception. Affect differs from emotion, but is important for action and interaction, and supplies the grounds from which we attribute emotions when we do.

In the final section, I further motivate this distinction by presenting evidence that affect perception and emotion perception come apart. I argue that the valence of others' affective states may be perceived in facial expressions by drawing on cross-cultural, psychopathological, and developmental evidence and arguments. If infant and cross-cultural abilities identify valence rather than emotion, that shows it is valence that is perceived rather than emotions per se. I take it that further research would be needed to demonstrate this capacity definitively but given that valence is a component of emotion expression and affective expressions, this is an interesting result and may also explain some of the differences between object recognition and emotion recognition sketched above.

Pacherie, Elisabeth (Institut Jean Nicod, CNRS/EHESS/ENS/PSL University)

The joys and sorrows of effort

On cost-benefit models of effort, the subjective experience of effort is the conscious output of sub-personal mechanisms designed to continuously compute the costs and benefits of engaging in the current task. Different cost/benefit models hold different views on what the currency of the incurred costs is (e.g., intrinsic costs (Kool and Botvinick, 2014), energetic costs (Gailliot & Baumeister 2007), or opportunity costs (Kurzban et al. 2013)) and on the exact form these cost/benefit computations take. Yet they all take the subjective experience of effort to be an intrinsically aversive experience, the magnitude of its negative valence corresponding to the cost computed relative to the benefit expected.

There is evidence, however, that people do not always try to avoid effort. They can also seek and value effort and effort exertion can feel enjoyable rather than unpleasant. This the so-called paradox of effort (Inzlicht et et al., 2018).

I will discuss two strategies for dealing with this paradox. The first strategy involves maintaining a cost-benefit approach to effort and the claim that effort is intrinsically aversive, but considering that expected additional rewards outweigh the costs of effortful performance. The second strategy involves renouncing the claim that effort is intrinsically aversive and showing that, under certain conditions, effort can be rewarding in itself.

According to the first strategy, for effort not to feel aversive, some counterbalancing incentive is needed, e.g., some expected future reward must outweigh the costs of effortful performance. This may involve incorporating into a cost-benefit computation not just the reward associated with the attainment of one's immediate goal but also the reward associated with the attainment of more distal goals, the satisfaction of which this immediate goal may contribute to. For instance, winning a difficult point in tennis may at a given moment in the competition be winning a game and securing a break, or winning a set, or even winning the match itself. Effort exertion may also yield social rewards, serving as a signal of one's commitment to one's goals and of one's value as a cooperative partner (Celniker et al. 2022). In addition, if increased effort consistently yields more rewarding outcomes, the sensation of effort itself may develop a conditioned association with positive reinforcers, counterbalancing its intrinsic aversiveness, a phenomenon known as learned industriousness (Eisenberger, 1992).

The second strategy constitutes a more radical departure from classical cost-benefit models of effort, holding that effort can be experienced as valuable or rewarding in its own right rather than merely because of its association with some external reward. Inzlicht et al. (2018) consider evidence that people sometimes derive value from effort itself, independently of any tangible products. Examples include phenomena such as need for cognition, that is a person's tendency to engage in and enjoy effortful cognitive activities for their own sake (Cacioppo et al., 1996), endurance sports where practitioners report valuing their sport precisely because it demands a great deal of effort (Loewenstein, 1999), or states of flow, where people derive pleasure from performing effortful activities in which they are fully immersed (Csikszentmihalyi, 1994).

I will argue that the second strategy for dealing with the effort paradox is more promising than the first. The first strategy assumes an instrumental conception of action, according to which we care for goals and the benefits they yield and pursue the means for the sake of the end. In contrast, the second strategy allows for what Nguyen (2020), in his book on agency in games, calls motivational inversion, cases where we pursue an end for the sake of the means, as when in some games we do not value a goal for its own sake but take it up for the sake of the activity of struggling for it.

Doing away with a strictly instrumental conception of action and allowing that we can pursue an end for the sake of the means as well as pursuing the means for the sake of an end allows for a richer and more nuanced view of the experience of effort, where the experience of effort can have multiple sources and multiple facets, reflecting valuation along a number of dimensions, linked not just to cost/benefit computations and social signaling, but also to the exercise of our agentive capacities (e.g., efficacy, mastery, self-control) and the pleasure we may derive from that exercise.

Pani, Silvana (University Assistant (Doctoral Fellow))

General and specific images: a distinction for the intention-motor interface and its implications for imagery construction

The idea that intentions are propositional states figuring in practical reasoning is a traditional platitude (e.g., Bratman 1987). The idea that motor representations, qua immediate antecedents of actions, are non-propositional in nature is a more recent view, though already widely accepted (e.g., Jeannerod 2006). The question of how differently formatted contents, that is, contents that are propositionally and motorically formatted come together towards the realization of some action goal has been dubbed the "Interface challenge" by Butterfill and Sinigaglia (2014). In this paper, I spell out what I take to be the two issues at the heart of the challenge: 1) the content determination question (Burnston 2017; Mylopoulos and Pacherie 2019), namely, the problem of how general or distal intentions acquire the specific contents that they do at the action implementation levels; 2) the format coordination question, namely, how coordination between differently formatted contents is to be explained.

My account of the interface answers both questions by introducing a largely underexplored distinction in imagination studies, that is, the distinction between general and specific images. Evidence in support of said distinction comes from different empirical venues and converges on the recognition of mental images as presenting various degrees of richness of detail. Coordination between images of different degrees of specificity is realized by a procedure of filling-in of details of portions or features of images. Similar to digital maps, images can be zoomed in and out depending on the function they are constructed to serve and the stage of action they are to guide.

When images are too general, they are unlikely to have any adequate verbal counterpart. As a matter of fact, retrieving general images does not necessarily require conceptual knowledge. Likewise, when images are highly specific, such as when they represent sensorimotor details relevant to motor tasks, they fail to have a corresponding verbal equivalent too. There is indeed no explicit evidence in computational models for motor control that internal representations are linguistic structures (see, for example, Wolpert and Ghahramani 2000). In all other cases, propositional components might be either integral to one's imaginings (see, e.g., Langland-Hassan 2015, 2020), or, as I will contend, might constitute sentential descriptions of map-like representations.

The current proposal is orthogonal to a few recent solutions to the interface challenge, especially Fridland (2021), where general intentions specify the overall goal or outcome of actions, while practical intentions specify the means to achieve that end. I contend that intentions often take the form of images, with general images representing outcomes and specific images representing means or aspects of outcomes. If it's true that intentions – contrary to the longstanding platitude – can live a double life in that they can somehow accommodate the propositional as well as the non-propositional format (Shepherd 2019), so can some (albeit not all) mental images leading up to overt action, thereby offering a solution to the challenge that has the potential to minimize the points of format interlock.

References

Bratman, Michael. "Intention, plans, and practical reason." (1987).

Burnston, Daniel C. "Interface problems in the explanation of action." *PhilosophicalExplorations* 20.2 (2017): 242-258.

Butterfill, Stephen A., and Corrado Sinigaglia. "Intention and motor representation inpurposive action." *Philosophy and Phenomenological Research* 88.1 (2014): 119-145.

Fridland, Ellen. "Intention at the interface." *Review of Philosophy and Psychology* 12.3(2021): 481-505.

Jeannerod, Marc. *Motor cognition: What actions tell the self*. Vol. 42. OUP Oxford, 2006.

Langland-Hassan, Peter. "Imaginative attitudes." *Philosophy and PhenomenologicalResearch* 90.3 (2015): 664-686.

Langland-Hassan, Peter. *Explaining imagination*. Oxford University Press, 2020.

Mylopoulos, Myrto, and Elisabeth Pacherie. "Intentions: The dynamic hierarchical modelrevisited." *Wiley Interdisciplinary Reviews: Cognitive Science* 10.2 (2019): e1481.

Shepherd, Joshua. "Skilled Action and the Double Life of Intention." *Philosophy andphenomenological research* 98.2 (2019): 286-305.

Wolpert, Daniel M., and Zoubin Ghahramani. "Computational principles of movement neuroscience." *Nature neuroscience* 3.11 (2000): 1212-1217.

Patronnikov, Ilia (University of Barcelona)

Implicit bias and the problem of inconsistent beliefs

The aim of my paper is to solve a certain problem that emerges from research on implicit bias. In a nutshell, the problem amounts to the following: experimental evidence suggests that beliefs that people have about different social groups are plainly inconsistent: one might believe that a social group is F, and at the same time believe that it is not F. If this picture, according to which a person has plainly inconsistent beliefs, is correct, it would render people extremely irrational, which they presumably are not. Some might argue that this provides a reason to reject the picture. This is not the strategy that I would suggest. Instead, I will argue that the picture doesn't necessarily entail extreme irrationality. I will discuss several ways to square the idea that people have plainly inconsistent beliefs with the idea that they are not extremely irrational. The best way to do so is to combine inferentialism about belief with the observation that the knowledge about the inconsistent beliefs in question is subject to motivated reasoning.

Pelland, Jean-Charles (University of Bergen)

Representation vs affordance: the case of number

If traditional computational psychology can be said to center around the notion of representation, the theoretical cornerstone of Gibson's ecological conception of the mind is the notion of affordance, which he introduced to capture the opportunities for action given to an organism by its environment. While Gibson's description of an affordance as "neither an objective property nor a subjective property; or it is both if you like" (1979, 121) may appear confusing, dispositionalist approaches to affordances (e.g. Greeno 1994; Scarantino 2003) have helped anchor this notion in more palatable and familiar vocabulary. Under such dispositionalist approaches, it is relatively straightforward to frame classic examples of affordances such as the 'grabability' of a cup or the 'mailability' of a mailbox in terms of the abilities of the organism and the objective properties of the objects in its environment.

However, while dispositionalist approaches to affordances may fare well with physical objects, our interaction with abstract objects remains an important challenge for affordance-based approaches to the mind. For example, explaining what numbers are in terms of affordances is particularly challenging for both ecological psychology and for the enactivist schools of thought that heavily rely on it (e.g. Hutto 2019; Gallagher 2017). The problem here is due to the abstract character of numbers: contrary to grabable cups and mailable mailboxes, numbers do not inhabit our physical environment. This means that it is far from obvious to find any features of the environment that could explain the enumerability of collections of objects in a way similar to how the grabability of a cup, say, can be explained in terms of the size of human hands and the shape of cups.

Those wishing to frame numerical cognition under a Gibsonian banner have two main options: explaining how direct perception of numbers might work, or explaining how we obtain numerical abilities by 'scaling up' from other, more basic affordances. In this talk, I argue that while some progress has been made on both these fronts, classical cognitivist approaches to cognition fare better than ecological alternatives in capturing numerical cognition for two main reasons: first, their ability to interpret and confirm empirical data, and, second, their ability to explain what makes certain cognitive processes specifically about numbers.

To make my case, I start by discussing problems with proposals made by philosophers of mathematics who attempted to frame mathematics in broadly Gibsonian terms over the years (e.g. Kitcher 1984; Maddy 1990; Kerkhove & Myin 2002). In a nutshell, I argue that these suffer from both conceptual problems and lack of empirical validity.

Regarding empirical validity, I assess whether affordance-based approaches can make sense of data concerning i) the so-called Approximate Number System (Dehaene 2011), ii) well-known developmental

milestones in the development of numerical abilities (Carey 2009) and iii) behaviour of subjects violation-of-expectation studies used to explore potential building blocks of our numerical abilities (Ginnobili & Olmos 2021). I argue that in all cases, representation-based accounts (e.g. Dehaene 2011; Carey 2009) fare better in explaining and respecting the data and methods.

As for the conceptual problem, it starts from the simple observation that not all animals can enumerate things. On the contrary, despite recent sensationalist claims that bees can categorize numerosity (Howard et al. 2022), only a small portion of humans have become truly numerate, in the sense of understanding what numbers are and being able to tell two arbitrary quantities apart. Given that only enculturated humans have managed to develop numerical abilities, it is important to explain what sort of special abilities, if any, are supposed to allow us to detect numerical affordances, and what sort of properties of the environment, if any, are supposed to afford counting and other numerical tasks. While classical cognitivist approaches to the mind can rely on representation-friendly interpretations of the data concerning potential cognitive systems recruited in counting and other arithmetical tasks (e.g. Cohen Kadosh & Dowker 2015) to characterize such abilities, I argue that none of the proposed affordance-based alternatives have managed to single out what would make an affordance number-specific. I argue that while Jones' (2018) proposal to see numbers as affordances for sequential attention does seem to improve upon previous attempts, the myriad non-numerical cases of sequential attention (e.g. choosing between two options by chanting "eeny-meeny-miny-moe") make this at best a necessary condition for numerical affordances.

After having discussed the possibility of direct perception of numbers, I turn my attention to attempts to scale up to numerical abilities from more basic affordances as illustrated by recent enactivist proposals (Hutto 2019; Gallagher 2017). Here, I show that not only the same conceptual and empirical worries apply as well, but that some versions of enactivism (especially the self-proclaimed 'radical' kind) seem to go against well-established empirical data regarding Weber effects in how we process numerical information presented in symbolic format (Dehaene 1996) by claiming that "privileging of the neural as an ultimate source of a key element of mathematical understanding" (Hutto 2019, 832) is problematic.

The upshot of the critical part of the talk is that what is missing in affordance-based alternatives is a difference-maker (Clark 1998) between numerical and non-numerical actions. I end the talk by offering a sketch of a solution to this problem by developing a novel account of mental affordances where individuation plays a central role. While McLelland (2020) has recently attempted to explain our numerical abilities in terms of mental affordances, his account faces the same problem as other affordance-based accounts due to its framing mental affordances in terms of mental actions on the physical environment, where, I claim, there are no features differentiate the numerical and non-numerical. In contrast, I offer an account of mental affordances in which mental actions take place in relation to a mental environment. This allows me to explain what numbers are by appealing to our ability to attend to individuated experiences of quantificational action, in line with Wu's (2014) account of attention as selection for action.

Pokropski, Marek (University of Warsaw, Faculty of Philosophy) and Piotr Suffczyński (University of Warsaw, Faculty of Physics)

First-person constraints on explanations in neuroscience. The case of migraine aura models

The paper focuses on the notion of explanation in cognitive neuroscience. We argue that mechanistic approach to explanation applied in cognitive neuroscience can be constrained by first-person studies of mental phenomena. Our argumentation is based on a case study of explanation of migraine with aura.

In the last two decades, a shift from cognitive science to cognitive neuroscience has taken place (e.g., Boone and Piccinini, 2016). The development of brain studies changed the way we think about mental phenomena and their explanation. In particular, the neomechanistic model of explanation became dominant and widely applied in cognitive neuroscience (e.g., Craver, 2007). According to this model, an explanation of a target phenomenon relies upon describing a causal mechanism which produces it.

Importantly, recent discussions about the relation between the mechanistic and dynamical framework show that the opposition between these frameworks is only apparent and that to some extent they are complementary (e.g., Kaplan and Bechtel, 2011). In other words, dynamical models which describe the target system state space can be supplemented with mechanistic details showing how the system's studied behavior is produced by certain mechanisms. Thus, according to Bechtel and Abrahamsen (2010), we may speak of a hybrid dynamic-mechanistic (DM) explanation.

Recently, it was also argued that the mechanistic framework is a road towards integration of cognitive neuroscience (e.g., Craver, 2007; Miłkowski 2016). The idea of mechanistic integration relies on integration of constraints that various research fields provides limiting the space of possible explanatory models. The key notion here is that of constraints. The success of an explanation depends on various constraints provided by different research fields that limit the set of possibilities and thus contribute to specification of the searched for mechanism.

Typically, in neuroscience, constraints concern structural or functional properties of the mechanism responsible for the target phenomenon and are formulated in research fields in which third-person methods of investigation are utilized. For example, neuronal behavior can be investigated in one research field focusing on the size and localization of neurons, whereas another field investigates the biophysical properties of neurons, such as membrane capacitance. An important type of constraints are dynamical ones, i.e., constraints which concern a change of behavior of a target system or its components in time. They are important because they indicate key properties of complex system behavior such as stable and unstable states, or that a stable state is oscillatory. Such information about dynamic properties of the studied behavior constrains the space of possible mechanisms to those which actually may produce the behavior, e.g., oscillatory mechanisms.

Not much attention has been given to constraints that could be provided by the study of first-person experience, which in the case of multifaceted mental phenomena are of key importance. Think, for example, about mental maladies, such as depression and schizophrenia, different types of addictive disorders, or visual aura experiences which accompany migraine headaches.

In this article, a general question we consider is whether first-person descriptions can constrain mechanistic explanations of mental phenomena. Our answer is positive. We argue that a dynamic-mechanistic framework is able to address first-person aspects of mental phenomena. Furthermore, in some cases first-person constraints may be an essential contribution to explanatory models. Our argumentation is based on a case study of migraine with aura.

Migraine headaches with auras are a common condition which occurs in 8% of the general population (Kirchmann, 2006). The aura is a subjective phenomenon which often accompanies headaches in migraine onsets. Although the phenomenology of aura experiences is diverse, including sensory feelings progressing through the body and aphasic symptoms, the most common type of aura is visual. The visual aura experience had been illustrated by migraineurs over the years, and these drawings were key in finding an mechanism which is responsible for the aura phenomenon i.e. the cortical spreading depression (CSD) (e.g., Schott, 2007). As we argue, in the case of migraine, first-person investigations have not only given us a description of the visual aura and better understanding of migraine, but provided key dynamical constraints on mechanistic models.

Taking into account the history of the development of migraine explanations, first-person insights about visual aura propagation in the visual field and the shape of aura percepts were essential contributions to the CSD hypothesis which underlies current explanatory models. The first-person information delivered dynamical and structural constraints on explanatory models. First, the shape and trajectory of the aura suggest that excitability of the cortex leading to CSD may be within the weak excitability regime. Second, the percept's zig-zag pattern followed by a scotoma indicated neuronal structures in the visual cortex, such as orientation columns in the V1 cortical area. Models informed by these first-person insight and related hypotheses (e.g. Dahlem and Müller 2003) were able to recreate spatial and dynamic properties of the aura's percept and make successful predictions about its progression.

Porat, Roy (Brown University / Tel Aviv University)

Semantic analysis of lexical synesthesia: What can it teach us about abstractness, Embodied Cognition, and the processing of proper names?

How do we process concepts such as "CAT"? And what about "FREEDOM" or "JOHN"? Theories of embodied cognition, which reject traditional views of cognition as computation on amodal symbols, suggest that the representation and processing of concepts are fundamentally grounded in perceivable experience [Barsalou 2008]. Abstract concepts, however, pose a particular problem for this model: while concrete words like "cat" may involve the internal simulation of the perceptual, motor and emotional experiences linked to their referents (seeing a cat, caressing it, hearing it meowing and so on), abstract words like "freedom" or "importance" do not have any perceivable referents by which they can be conceptualized [Borghi et al 2017]. In order to cope with this problem, some proponents of the embodied view suggest that the human mind compensates for this lack of perceivable features by recruiting replacements from more concrete, sensory domains, like temperature, weight, or spatial relationships; these mappings, as suggested, provide a straightforward link between the physical and the representational, enabling people to comprehend abstract concepts by grounding them in concrete sensorimotor experience [Lakoff and Johnson 1980, 1999; Gibbs 2006].

So far, this model has been supported mainly by two lines of evidence: first, analysis of recurring linguistic patterns – commonly known as Conceptual Metaphors – which are assumed to represent verbal manifestations of these hidden conceptual structures; and second, dozens of clever priming experiments, which revealed implicit associations between hypothetically-related domains such as affection and warmth ("she is a warm person"), importance and physical weight ("light/heavy matter"), moral purity and cleanness, and many more [Williams & Bargh, 2008; Jostmann et al 2009; for review, see Shen & Porat 2018]. Convincing as they may be, these studies do leave a theoretical "gap", as they rely exclusively on commonly shared associations, and thus do not account for potential mappings that would not fall neatly under the familiar metaphorical schemes. But what if conceptual grounding can also be idiosyncratic, namely, one that derives entirely from an individual set of associations, like "importance" and warmth, or "freedom" and the color blue? While this new kind of evidence for the embodied view may remain unnoticed using the standard methodologies, we suggest that the study of cross-modal interactions – or more specifically, semantic analysis of a specific kind of lexical synesthesia – shows evidence of such relations.

Since synesthesia is almost unanimously regarded by the scientific community as a strictly perceptual phenomenon – according to the most widely-accepted model, the result of enhanced cortical connectivity between brain regions responsible for processing different sensory modalities – semantic or linguistic analysis of its occurrences has been extremely rare, and was used mainly in anecdotal contexts (e.g., as determiners of specific synesthetic matchers, like the word "cinema" and the flavor of cinnamon rolls). Despite this common view, however, the most common synesthetic inducers are conceptual artifacts like weekdays, months, number and letters, a fact that casts doubt on the standard definition of synesthesia as a purely perceptual phenomenon. In the current study we have conducted an analysis of the semantic categories in cases of lexical synesthesia, a form of synesthesia where words act as inducers, usually in the context of words-flavors ("lexical-gustatory" synesthesia). The results were conclusive: abstract words (e.g., "democracy", "freedom") have a significantly greater chance of becoming synesthetic inducers compared to concrete words ("chair", "umbrella"). Furthermore, of the lexical synesthetes who took part in the advanced stages of the study, half demonstrated an almost absolute "abstract" synesthesia (with only abstract words and names serving as synesthetic inducers, while concrete words induce no synesthetic experience whatsoever), implying that this might in fact be the most common type of lexical synesthesia, a claim that due to the absence of semantic analysis in previous studies has not been examined before.

In my talk I will discuss the implications of the above findings, first and foremost the intriguing possibility that "abstract synesthesia" forms not only the first documented case of semantic-based synesthesia but rather a specific form of idiosyncratic grounding, and thus filling the potential "gap" in (and providing a novel kind of support of) the embodied theory of representation. This form of synesthesia, according to my suggestion, is not a perceptual but rather an inherently conceptual phenomenon triggered by semantic factors: the need to "compensate" for the lack of concrete features during the processing and acquisition

of abstract words. I will further discuss the possibility that this model is not limited to the specific kind of lexical synesthesia examined, but rather that the most common types of the condition – colored weekdays, months, letters and numbers – all form particular cases of abstract synesthesia. This suggestion, alongside several other phenomena which were discovered in the study such as multi-concurrent synesthesia and abstract synesthesia between words and images, challenge the common distinction between "real" synesthesia and semantic association, redefining the relationship between the synesthetic inducer and induced experience as somewhat parallel to the one of signifier and signified. In addition, I will expand on one specific category that was proven to be a particularly strong type of inducers, the one of proper names, which, as I suggest, derives from the tension between the hyper abstractness of names on the one hand and their alleged concreteness on the other ("John" implies on the existence of a concrete object, a person that carries that name, but provides basically no information about its referent). As I will show, this finding bears on the long old debate among philosophers of language regarding the meaning of names, providing strong support of the Millian theory of reference [Mill 1843; Kripke 1980].

References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, *143*(3), 263.

Gibbs, R. W. J. (2006). Embodiment and cognitive science. New York, NY: Cambridge University Press

Jostmann, N. B., Lakens, D., & Schubert, T. W. (2009). Weight as an embodiment of importance. *Psychological science*, *20*(9), 1169–1174.

Kripke, S. A. (1980). *Naming and necessity.* Harvard University Press.

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago, IL: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenge to western thought. New York, NY: Basic books.

Mill, John Stuart, 1843, *A System of Logic*, London: Parker.

Ramachandran V S, Hubbard E M, 2001 "Synaesthesia: A window into perception thought and language" Journal of Consciousness Studies 8 3 ^ 34

Shen, Y., & Porat, R. (2018). "Metaphorical Directionality: The Role of Language" in Hampe, B. (Ed.). *Metaphor: Embodied cognition and discourse*. Cambridge University Press.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., Scott, K., Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. Perception, 35, 1024–1033.

Ward, J., & Simner, J. (2003). Lexical-gustatory synaesthesia: linguistic and conceptual factors. *Cognition*, *89*(3), 237–261.

Ward, J., & Simner, J. (2022). How do Different Types of Synesthesia Cluster Together? Implications for Causal Mechanisms. *Perception*, *51*(2), 91–113.

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322* (5901), 606–607.

Zhong, C. B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, *313* (5792), 1451–1452.

Quint, Andrea (University of Heidelberg), Phuc Nguyen (University of Erlangen) and Nora Heinzelmann (University of Erlangen)

The ethics and moral psychology of vaccination: evidence from a cross-cultural study

The COVID-19 pandemic has propelled academic research into the ethics of vaccination to the forefront of public discussion. One side of the debate argues that not getting vaccinated is unethical because it imposes serious risks on others; punishing free-riders may, from this perspective, be legitimate to establish the common good of herd immunity (Choy & Yong, 2022). The other side of the debate objects that a decision about vaccination is a personal and prudential one, and that contributing to herd immunity is

at best supererogatory but not ethically required; on this view, a vaccine mandate or penalties against the unvaccinated violates individual liberty and autonomy (cf. Braddock, 2013).

Increasingly, the debate about vaccination has become moralised beyond academic circles, which in turn may negatively affect public health (Rosenfeld, 2022). In many countries, vaccination rates for preventable diseases are now far below the ones needed for herd immunity, including COVID-19, measles, influenza, and some forms of cancer (World Health Organization [WHO], 2022).

Our research seeks to addresses the issue by first gaining an insight into the moral psychology of decision-making about vaccination and then developing suggestions into how this process of decision-making could be improved. Here, we report preliminary findings from a pre-registered, cross-cultural study in representative samples from the US, the UK, and Germany (N=1200). The goal of this study was to identify factors that determine decision-making about vaccinations, using the case of vaccination of children against cancer as an example. We collected survey data from parents on the online platforms Qualtrics and Prolifico.

In particular, we examined to what extent, if any, participants view the decision as an ethically salient one or feel morally pressured to have their children vaccinated (or not). In line with empirical findings, we also hypothesized that believing to know more about a vaccine increases an agent's confidence in their decision, with those who actually do know more being more likely to actually get vaccinated (Caddedu et al., 2021, Ebrahimi et al., 2021, Abu Hammour et al., 2021). We also tested for the role of trust in doctors, science, and medical institutions.

Preliminary findings indicate that participants do view the decision about vaccination as an ethically salient one. For example, the ethically better participants rate the vaccination, the greater is their willingness to have their children vaccinated ($R2 = .39$, $p < .01$). Also, the stronger parents agree that they are ethically required to vaccinate their child for other people's benefit, the higher is their willingness to vaccinate their children ($R^2 = .17$, $p < .01$).

Trust in doctors, science and medical institutions seems to strongly determine parental decision making about vaccination. A multiple regression analysis including interaction effects showed a connection between a higher trust in physicians, science and health institutions and a lower negative association between risk perception and willingness to have one's own children vaccinated against HPV (adj. $R^2 = .43$, $p < .01$). The individual interaction effects are shown Figure 1, below. In other words, the more participants trust doctors, science in general, and the health institutions of their country, the more willing they are to have their children vaccinated.

Strikingly, this effect was much stronger and more significant than the expected effect of knowledge. In particular, we did not find evidence that either knowledge about the vaccine or the disease it targets correlates with the willingness to get vaccinated. We did find a link between ethical ratings of the vaccine and both knowledge about it ($R^2 = .01$, $p = .02$) and about the cancer it targets ($R^2 = .01$, $p = .04$).

Our findings contribute to a better understanding of the moral psychology of decision-making

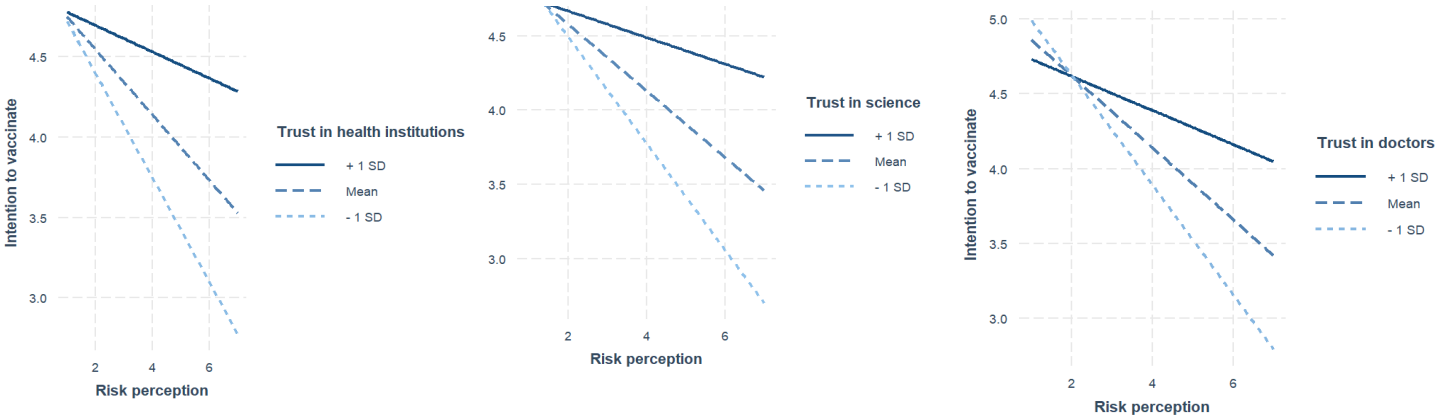


Figure 1. Vaccination and trust. *Willingness to get vaccinated is strongly correlated with trust in physicians (a), science in general (b), and trust in health institutions (c). Intention to get vaccinated was measured on a Likert scale ranging from 1 (no intention to vaccinate) to 5 (full intention to vaccinate), adapted from Caso et al. 2021. Trust was measured as a composite variable from 1 (lowest) to 7 (highest). We used the trust in the medical profession scale by Dungan et al. 2005 in (a), the trust in science and scientists inventory from Nadelson et al. 2014 and Krüger et al. 2022 in (b), and the trust in healthcare institutions measure by Caso et al. 2021 in (c).*

about vaccination and may have ethical implications. In particular, they indicate how patients, medical professionals, healthcare providers and institutions may shape the decision-making process. For one thing, as decision-makers regard vaccination as ethically salient, ethical arguments for or against vaccination may be given and considered. Moreover, establishing a trusting relationship between decision-makers and physicians, the scientific community, and health institutions may beneficially affect vaccination rates. In contrast, information campaigns and education may have a lesser impact than previously expected. In our talk, we plan to address these aspects and others in further detail, such as cross-cultural comparisons that may explain diverging rates of vaccination between the countries studied.

Bibliography

Abu Hammour, K., Abu Farha, R., Manaseer, Q., & Al-Manaseer, B. (2022). Factors affecting the public's knowledge about COVID-19 vaccines and the influence of knowledge on their decision to get vaccinated. *Journal of the American Pharmacists Association*, *62*(1), 309–316.

Ahmad, M., & Majid, U. (2022). The value of information: a qualitative analysis of how trust in information sources influences the decision to vaccinate in parents. *Journal of Risk Research*, *25*(11–12), 1444–1457.

Braddock, M. (2013b). Defusing the Demandingness Objection: Unreliable Intuitions. *Journal of Social Philosophy*, *44*(2), 169–191.

Cadeddu, C., Castagna, C., Sapienza, M., Lanza, T. E., Messina, R., Chiavarini, M., Ricciardi, W., & de Waure, C. (2021). Understanding the determinants of vaccine hesitancy and vaccine confidence among adolescents: a systematic review. *Human Vaccines & Immunotherapeutics*, *17*(11), 4470–4486.

Caso, D., Capasso, M., Fabbricatore, R., & Conner, M. (2022). Understanding the psychosocial determinants of Italian parents' intentions not to vaccinate their children: an extended theory of planned behaviour model. *Psychology & health*, 37(9), 1111-1131.

Choy, B. K. C., & Yong, J. C. (2022). Free Rider Problem During the Covid-19 Pandemic. *Encyclopedia of Business and Professional Ethics*, 1–6. https://doi.org/10.1007/978-3-319-23514-1_1313-1

Hall, M. A., Camacho, F., Dugan, E., & Balkrishnan, R. (2002). Trust in the medical profession: conceptual and measurement issues. *Health services research*, 37(5), 1419-1439.

Ebrahimi, O. V., Johnson, M. S., Ebling, S., Amundsen, O. M., Halsøy, Y., Hoffart, A., Skjerdingstad, N., & Johnson, S. U. (2021). Risk, Trust, and Flawed Assumptions: Vaccine Hesitancy During the COVID-19 Pandemic. *Frontiers in Public Health*, *9*.

Krüger, J. T., Höffler, T. N., & Parchmann, I. (2022). Trust in science and scientists among secondary school students in two out-of-school learning activities. *International Journal of Science Education, Part B*, *12*(2), 111-125.

Nadelson, L., Jorcyk, C., Yang, D., Jarratt Smith, M., Matson, S., Cornell, K., & Husting, V. (2014). I just don't trust them: the development and validation of an assessment instrument to measure trust in science and scientists. *School Science and Mathematics*, 114(2), 76-86.

Rosenfeld, D. L., & Tomiyama, A. J. (2022). Jab my arm, not my morality: Perceived moral reproach as a barrier to COVID-19 vaccine uptake. Social Science & Medicine, 294, 114699.

World Health Organization. *Immunization coverage*. (2022, July 14). https://www.who.int/news-room/fact-sheets/detail/immunization-coverage

Raia, Renato (University of Milan)

Dissolving the debate on the format of concepts

A key debate in the philosophy of cognitive science concerns the format in which concepts are represented. This debate has been structured by the contrast between two main positions: modalism (in the literature, also "concept empiricism" [1] or "neo-empiricism" [2]) and amodalism. Modalists, whose position can be traced back to the British empiricists [3], hold that conceptual representations share the same format as sensory, motor, and affective representations, such that each concept is the result of offline reenactments of the perceptual experience one has had with the external entities represented by the concept [4]. In contrast, according to amodalist perspectives, which have been the default view in classical cognitive science in recent decades, conceptual systems have their own representational format which abstracts from the sensory modality of the stimuli, and which can be described as language-like (although they do not necessarily constitute a language) [5].

Over the course of the past two decades, modalists and amodalists have sought to support their respective positions by appealing to a broad range of empirical evidence as well as theoretical arguments, which I evaluate in this paper. I argue that neither the evidence nor the theoretical arguments unambiguously support either position, and diagnose the underlying conceptual ambiguities in both theoretical positions, which currently stand in the way of progress in this debate.

I begin by evaluating the empirical evidence. The main empirical support for a modalist conception comes from the neural reuse evidence: neuroimaging studies confirm the activation of sensorimotor areas of the brain during the processing of sensory- or motor-related concepts (words) ([6], [7], [8], [9], [10], [11], [12], [13], [14]). Furthermore, there is evidence of a cost in terms of the efficiency of conceptual processing when the processed word switches from referring to one modality to another. This modality switching cost has been used to argue that conceptual processing depends upon perceptual modalities [15].

On the other hand, the main evidence for amodalism comes from a neuropathological condition called semantic dementia (SD), which leads to a progressive loss of conceptual knowledge as result of a focal damage to the anterior temporal lobe (ATL). Patients with this condition lose knowledge of specific concepts across all sensory modalities, while preserving access to other concepts through the same modalities. This has led some researchers to argue that concepts are represented in an amodal fashion ([5], [16]).

Modalists have replied to the findings from SD by stating that actual conceptual processing happens in the sensorimotor areas and that the ATL just functions as a "supramodal" area, joining modal concepts into multimodal ones [17]. Amodalists, in contrast, have reinterpreted the data about neural reuse by stating that actual conceptual processing happens in amodal areas and that the concurrent activation in the sensorimotor systems is just a way of "offloading" processing to perceptual areas as a sometimes-useful heuristic [18].

I argue that these reinterpretations of the empirical data are theoretically weak: if there is activation in both the ATL and connected areas and in the perceptuomotor systems there is no clear way to ascertain whether either process is directly involved in actual conceptual processing. Why should we believe the reuse of the neural base of perceptual systems is an instance of "offloading" as opposed to being a proper part of conceptual processing? On the other hand, the data coming from SD cannot be dismissed by stating that the ATL performs some further function while real concepts are stored in the perceptual areas. The data shows the activation of both clearly perceptuomotor and clearly amodal areas during conceptual processing, so one should conclude that both areas are involved in concept formation and that the neural location criterion is useless for adjudicating the debate [19].

In sum, the empirical evidence so far obtained in this context provides valuable constraints on theorizing about the format of conceptual representations, but it does not resolve the debate between modalism and amodalism. Indeed, before conducting further empirical investigation with a view to adjudicating this debate, it would be crucial to spell out each theoretical position in greater detail, and to derive specific testable predictions which differentiate between the two positions in a principled manner.

I then turn to theoretical arguments. One argument in favor of amodalism comes from abstract concepts [20] and particularly from our number concepts [18]. These concepts seem hard for modalists to account for, especially because our innate faculty for numerosity estimations seems not to present a modality switching cost ([21], [22]). However, modalists can account for abstract and number concepts by stating that they are the result of an abstraction and convolution mechanism operating over modal representations ([23], [24]). Modalists have been challenged to provide an empirical basis for discriminating between such multimodal system and amodal ones [18]

In addition, I argue that there is an inherent ambiguity in the terms "multimodal" and "amodal" so that it is by principle impossible to establish whether modal representations which have gone through a process of abstraction can still be considered modal or amodal. This is the reason why modalists (and amodalists) cannot provide a criterion to establish whether high-level representations in the ATL, for example, count as evidence for amodalism or multimodality. In sum, the empirical evidence so far obtained in this context, and the theoretical arguments so fa far developed, do provide valuable constraints on theorizing about the format of conceptual representations, but do not resolve the debate between modalism and amodalism. Indeed, before conducting further empirical investigation with a view to adjudicating this debate, it would be crucial to spell out each theoretical position in greater detail, and to derive specific testable predictions which differentiate between the two positions in a principled manner.

Bibliography

Prinz J. J. (2002), *Furnishing the mind: Concepts and their perceptual basis*, MIT Press

Machery E. (2007), *Concept empiricism: a methodological critique*, Cognition, 104, 19–4

Barsalou L. W. (2008), *Grounded cognition*, Annu. Rev. Psychol., 59, 617–645

Barsalou L. W. (1999), *Perceptual symbol systems, Behavioral and brain sciences*, 22(4), 577–660

McCaffrey J. & Machery E. (2012), *Philosophical issues about concepts*, Wiley Interdisciplinary Reviews: Cognitive Science, 3(2), 265–279

González J., Barros-Loscertales A., Pulvermüller F., Meseguer V., Sanjuán A., Belloch V. & Ávila C. (2006), *Reading cinnamon activates olfactory brain regions*, Neuroimage, 32(2), 906–912

Kiefer M., Sim E. J., Herrnberger B., Grothe J., Hoenig K. (2008), *The sound of concepts: four markers for a link between auditory and conceptual brain systems*, J. Neurosci., 28, 12224–12230

Pulvermüller F. & Hauk O. (2006), *Category-specific conceptual processing of color and form in left fronto-temporal cortex*, Cerebral Cortex, 16(8), 1193–1201

Reuter K., Werning M., Kuchinke L. & Cosentino E. (2017), *Reading words hurts: the impact of pain sensitivity on people's ratings of pain-related words*, Lang. Cognition, 9(3), 553–567

Simmons W. K., Martin A. & Barsalou L. W. (2005), *Pictures of appetizing foods activate gustatory cortices for taste and reward*, Cerebral Cortex, 15, 1602–1608

Goldberg R. F., Perfetti C. A. & Schneider W. (2006), *Perceptual knowledge retrieval activates sensory brain regions*, Journal of Neuroscience, 26, 4917–4921

Hauk O., Johnsrude I. & Pulvermüller F. (2004), *Somatotopic representation of action words in human motor and premotor cortex*, Neuron, 41(2), 301–307

Hauk O. & Pulvermüller F. (2004), *Neurophysiological distinction of action words in the fronto-central cortex*, Human Brain Mapping, 21(3), 191–201

Pulvermüller F. (2005), *Brain mechanisms linking language and action*, Nature reviews neuroscience, 6(7), 576–582

Pecher D., Zeelenberg R., Barsalou L. W. (2003), *Verifying properties from different modalities for concepts produces switching costs*, Psychol. Sci., 14: 119–124

McCaffrey J. (2015), *Reconceiving conceptual vehicles: Lessons from semantic dementia*, Philosophical Psychology, 28(3), 337–354

Kiefer M. & Pulvermüller F. (2012*), Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions*, Cortex, 48(7), 805-825

Machery E. (2016), *The amodal brain and the offloading hypothesis*, Psychonomic bulletin & review, 23, 1090- 1095

Haimovici S. (2018), The modal—amodal distinction in the debate on conceptual format, Philosophies, 3(2), 7

Dove G. (2009), *Beyond perceptual symbols: A call for representational pluralism*, Cognition, 110(3), 412–431

Barth H., La Mont K., Lipton J., Dehaene S., Kanwisher N. & Spelke E. (2006), *Non-symbolic arithmetic in adults and young children*, Cognition, 98, 199–222

Izard V., Sann C., Spelke E. S., & Streri A. (2009), *Newborn infants perceive abstract numbers*, Proceedings of the National Academy of Sciences, 106, 10382–10385

Michel C. (2021), *Overcoming the modal/amodal dichotomy of concepts*, Phenomenology and the Cognitive Sciences, 20(4), 655–677

Jones M. (2016), *Number concepts for the concept empiricist*, Philosophical Psychology, 29:3, 334–348

Patterson K. & Ralph M. A. L. (2016), *The hub-and-spoke hypothesis of semantic memory*, in "Neurobiology of language" (pp. 765–775), Academic Press

Raja, Vicente (University of Murcia)

The Motifs of Radical Embodied Neuroscience

In this paper, I develop a methodological tool that allows for the classification of the different paradigms and approaches within the sciences of the mind: the notion of motif. All paradigms and approaches that have enjoyed some degree of success within psychology (e.g., functionalism, behaviorism, Gestalt) and within neurophysiology (e.g., cognitive neuroscience) have been up to give answers to the questions of five fundamental themes that emerged during the 19th century:

(a) How do we understand the relationships between, stimuli, sensations, and experience?

(b) How do we understand the localization vs. holism debate?

(c) What are the (neural) mechanisms of experience?
(d) What are the right methods to study those mechanisms?
(e) What is the scope of the approach?

To provide these answers, different paradigms and approaches in the sciences of the mind do not offer strict guidelines but open-ended motifs. As an illustration, the ways cognitive psychology and ecological psychology characterize stimuli or the role of the nervous system in perception are largely incompatible. On the one hand, cognitive psychologists appeal to relatively simple stimulation along with relatively sophisticated internal event that process the information available in stimuli to construct a representation of the world. On the other hand, ecological psychologists reject to patch-up perception-action loops with internal processing powers and take the information available in stimulation to be rich enough as to support complex behavior. However, what cognitive psychologist take for "stimulation" or "representation" is famously diverse, and the same can be said for what ecological psychologists take for "information" or "processing". It is not just the disagree among paradigms; they just not agree within paradigms either. And the reason of this is that whatever concepts are regarded as the core one within these characterizations, they are not as profound as to fully prescribe all the activities of these approaches. For this reason, it is best to understand them not as golden laws thoroughly enforced by a particular theory but as recurrent motifs of a musical piece. For example, the characterization of the theme of stimuli in terms of information is one of these motifs in cognitive psychology. Not all cognitive psychologists are explicitly appealing to Shannon information or using its information-theoretic formulation in their research, but Shannon information is a recurrent motif in the cognitive psychology literature. It appears often when cognitive psychologists talk about stimulation. Just like the motif of "the good"—one note here, one down, and back again—appears often in the soundtrack of The Lord of the Rings for the musical themes of good characters like Gandalf or the Fellowship. These motifs do not fully determine what happens with the theory or the story, but they provide a general scaffolding with regard to their main themes. These motifs tell the audience stimuli are generally characterized in terms of probabilistic information despite nuanced details of particular models, or that Frodo is good even if he behaves badly sometimes.

The idea of different paradigms dealing with the core themes of the sciences of the mind by using different motifs vertebrates this paper. Every theory of perception, action, and cognition may be seen as a soundtrack or a symphony composed by some core themes in which a handful of motifs are combined into a whole. If some motifs are combined in one particular way, you have cognitive psychology. If other motifs are at play, you have ecological psychology. Still other motifs lead to systems neuroscience. And, of course, a radical embodied neuroscience must have its own set of motifs for the core themes of the sciences of the mind. This talk will offer such set of motifs and will evaluate where they come from. To do so, I will take 3 consecutive steps.

In the first step, I will properly introduce the notion of motif and show how they can be used to categorize different paradigms and approaches in the sciences of the mind. A categorization based on motifs entails several benefits. For instance, it is detail enough as to allow for a clearcut distinction between different paradigms and approaches while it is not strict enough as to be unable to recognize the differences internal for particular paradigms and approaches. In the second one, I will introduce the motifs of what I call "the other psychology": a psychology that runs from William James' radical empiricism to ecological psychology (Chemero, 2009; Gibson, 1966, 1979) through phenomenology (Merleau-Ponty, 2012). I will show how these motifs of an eminently psychological approach to the mind can inform a neurophysiological one. Such neurophysiological approach is what I call radical embodied neuroscience. In the last step, I will propose and describe the motifs for such radical embodied neuroscience as a paradigm (or approach) within the sciences of the mind: the motif of complex stimulation, the motif of neural reuse (Anderson, 2014), the motif of resonance (Raja, 2018, 2019, 2021), the motif of the complexity methods, and the motif of full scope. This are the motifs able to deliver a coherent and detailed story that gives us a brain for the embodied mind.

References

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. The MIT Press.
Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, Mass. : MIT Press.
Gibson, J. J. (1966). *The senses considered as perceptual systems*. Greenwood Press.
Gibson, J. J. (1979). *The ecological approach to visual perception*. New York, NY : Psychology Press.
Merleau-Ponty, M. (2012). *Phenomenology of perception*. Routledge.

Raja, V. (2018). A Theory of Resonance: Towards an Ecological Cognitive Architecture. *Minds and Machines*, *28*(1), 29–51. https://doi.org/10.1007/s11023-017-9431-8
Raja, V. (2019). From metaphor to theory: The role of resonance in perceptual learning. *Adaptive Behavior*, 105971231985435. https://doi.org/10.1177/1059712319854350
Raja, V. (2021). Resonance and radical embodiment. *Synthese*, *199*(S1), 113–141. https://doi.org/10.1007/s11229-020-02610-6

Ranzenigo, Damiano (University of Konstanz)

Existential Self-identification

Aim of this paper is to present the concept of 'existential self-identification' (ESI, from now on), according to which some people come to consider some of their strong and intrinsic evaluative states as constitutive of how they conceive of their lives as meaningful. Examples of existentially self-identifying people are found among some fervent religious believers and political activists, who cannot conceive of a meaningful life outside of their respective ideologies, some queer-identifying people, who think that loving themselves as members of the LGBT+ community is necessary for their lives to be fulfilling, some 'Incels', 'furries', and other members of online communities, who adopt a specific self-understanding they deem as the only viable one to find purpose in life, some people deeply passionate about their professions, who are convinced they would be lost without their working routines, some people in parental, familial, and romantic relationships and close friendships, who cannot fathom a purposeful existence without their close ones, and many others.

Although phenomena of practical self-identification broadly understood have been the focus of works in philosophy of recent decades, following the pioneering works of Harry Frankfurt (1998, 1999) and Christine Korsgaard (1996, 2009), little attention has been devoted to the philosophical study of more specific forms of psychological self-identification, among which ESI is found. On the other hand, research on self-identification in more empirically oriented psychology has mainly focused on the content of identification practices within rather specific social groups, either taking for granted a shared understanding of the concept of identification or basing the interpretation of the findings on the theoretical scaffolding of social identity theory (see Burke and Stets, 2023 for an introduction). In recent decades, social psychologist Michael Hogg has also taken pains at developing his 'uncertainty-identity theory' (2007, 2021), according to which people would be keener to identify with a certain social group if they experience great self-uncertainty, but again, no new theory of identification has been advanced in support of these interesting correlations.

Working out a viable concept of ESI in a philosophical fashion allows for the more fine-tuned study of identification practices among at least some extremist individuals and social groups, queer-identifying people, members of online subcultures, and many others.

***

ESI is an eminently reflective phenomenon: the existentially self-identifying person is fully conscious of valuing, loving, or caring for a certain self-ascribed property. Although agents might ordinarily value, love, or care for many things without a conscious appreciation of so doing, there are no instances of subconscious ESI. This is also reflected in the fact that only at the reflective level one can conceive of a strong and intrinsic evaluative state as constitutive of the conception of one's life as meaningful.

Aside from monomaniac people, who greatly value one single thing in their lives, no strong and intrinsic evaluative state can be considered constitutive of a person's life's meaningfulness on its own, where for something A to be constitutive of something B is for A to be at once necessary for and part of B. Whilst strong and intrinsic evaluative states participate to the meaningfulness of a person's life, it is never the case that frustration of any such evaluative states disrupts a person's sense of meaning altogether because, for as long as a person values, loves, or cares for something else, she will find purpose, fulfilment, and meaning in her life.

Things change when it comes to existentially self-identifying agents. Such agents sincerely and consciously conceive of their life's meaningfulness as depending entirely on a particular evaluative state.

Of course, the belief that the thing they identify with is constitutive of their life's meaningfulness might be wrong. After all, they might realize that their life can go on reasonably well after losing the object of their evaluative state. However, they cannot be wrong about at least two things: (1) that they indeed strongly and intrinsically value the thing they existentially self-identify with; and (2) that they cannot consciously think about living a meaningful life without the object of their evaluation. People can be wrong about valuing things they actually do not care much about, but as soon as an agent thinks that an evaluative state is constitutive of his life's meaningfulness, he will do everything in his power to prevent its frustration. In displaying related cognitive, volitional, emotional, and behavioural dispositions, the agent cannot be wrong in strongly and intrinsically valuing the thing he existentially self-identifies with, as for (1). On the other hand, despite not being constitutive of the meaningfulness of a person's life, the strong and intrinsic evaluative state involved in ESI is constitutive of the conception of a person's life as meaningful, as for (2). This latter observation doesn't diminish the psychological impact that ESI has within an agent's psychic economy. Beside enhancing a person's willpower and self-certainty, in the most dramatic cases, ESI can also lead to diminished responsiveness to factual challenges, making some existentially self-identifying agents more vulnerable to conspiratorial or reality-denying ideological thinking.

***

ESI usually originates from the attempt at overcoming moments of existential distress, characterized by a sense of helplessness and loss of trust, familiarity, and comfort within one's environment – one need only think about climate activism in the face of the current climate crisis, queer people in rigidly heteronormative societies, right-wing and religious extremists feeling threatened by the increasing liberation and empowerment of historically disenfranchised groups, just to name a few. Coming to sincerely believe that one's life's meaningfulness depends on one single thing is not something one can bring about by an act of will, but rather something that happens to a person on a quest to regain her lost contact with the world by accessing some tangible bedrock for purpose and meaning in life. Nevertheless, the reflectiveness of ESI betrays the fact that without this active quest, a person would have never been able to spontaneously overcome her distress. Thus, ESI can be seen as a dangerous phenomenon, but also as a positive force to overcome difficult times.

Rappe, Sofiia (Ruhr University Bochum) and Markus Werning (Ruhr University Bochum)

Episodic memory and causal reasoning via counterfactual simulation: an approach based on trace minimalism and predictive processing

Over the past several years, there has been a shift in researchers' thinking about the functional role of episodic memory (Klein, 2013). Rather than focusing on how memory represents the past, recent literature often presents memory as ultimately dealing with the future – helping the organism to anticipate events and increase its adaptive success (Suddendorf & Corballis, 2007). Yet, in many cases, semantic and episodic memory can make available the same content (Klein, 2013) and potentially fill the same adaptive roles. For example, although episodic memory may serve as a source of vivid, experiential detail for future-oriented simulations (Buckner & Carroll, 2007; Schacter & Addis, 2007), the research shows that the contribution of episodic memory in the construction of (at least) autobiographical events diminishes in older adults, and the details are filled with the help of semantic, conceptual information (Addis et al. 2011). This raises a question: does episodic memory serve a unique adaptive role and has a unique evolutionary advantage?

We claim that episodic memory yields adaptive success because of its crucial role in causal reasoning. Often, it is assumed that predicting the future (and hence deciding which action to take) comes down to letting the internal causal model update "as if" the future events would unfold over time according to certain regularities stored in the semantic memory (Beck and Rafetseder 2019). If this were the case, we would only have to rely on accumulated statistical knowledge to decide amongst multiple hypothetical future scenarios depending on the action taken. However, basing decisions on statistical regularities might be misguided when the future is affected by exceptional singular events. Moreover, understanding the

causes of events and establishing causal laws for future decision-making in the first place cannot solely rely on the accumulation of statistical data, especially since such data is often sparse.

According to Lewis (1973), to causally reason whether event A caused event B, the subject needs to cognitively evaluate a diachronic counterfactual of the following type: (1) If A had not happened at t1, B would not have happened at t2. Such counterfactuals require episodic memory and mental time travel for their evaluation because they are true just in case (2) the closest to the actual world where A did not occur is also a world where B did not occur. However, mental time travel is a notorious problem for current associative network approaches used in AI and computational modeling in neuroscience. Such networks struggle with diachronic temporal reasoning because they do not explicitly represent change but rather update representations as new information comes along (Hoerl & McCormack, 2019). Second, they learn by altering the weights/strength of connections through accumulating data over time. This approach significantly differs from human learning, which, in some instances, only requires single examples and allows rapid, radical re-evaluation of the causal relationships between events.

Our paper presents a causal inference model based on the predictive processing framework of brain functioning (Clark, 2013, 2015; Hohwy, 2013; Friston 2005, 2010; Rao & Ballard, 1999) and minimal trace account of episodic memory (Werning, 2020). According to our model, because of its truth condition (2), evaluating counterfactuals of type (1) involves a) episodic memory to construct a scenario of the past (Cheng, Werning, Suddendorf, 2016), b) negation of the target event A in that scenario, and c) evaluation of the closeness relation between the newly attained non-A scenario and the remembered scenario. For example, to understand that the fact that your train was 30 minutes late caused you to miss the flight (rather than a 10-minute queue at the airport), you need to simulate the world as it existed when you expected the train to arrive according to the plan (a). Then, you need to simulate a world where the train did not get delayed by negating that event in the respective scenario (b). Finally, using semantic knowledge, you let the scenario unfold in time and ask yourself whether you would still have missed the flight or not (c).

Bibliography

Addis, D. R., Roberts, R. P., & Schacter, D. L. (2011). Age-related neural changes in autobiographical remembering and imagining. *Neuropsychologia*, *49*(13), 3656-3669.

Beck, S. R., & Rafetseder, E. (2019). Are counterfactuals in and about time?. *Behavioral and Brain Sciences*, *42*.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in cognitive sciences*, *11*(2), 49-57.

Cheng, S., Werning, M., & Suddendorf, T. (2016). Dissociating memory traces and scenario construction in mental time travel. *Neuroscience & Biobehavioral Reviews*, *60*, 82-89.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Hoerl, C., & McCormack, T. (2019). Thinking in and about time: A dual systems perspective on temporal cognition. *Behavioral and Brain Sciences*, *42*, e244.

Hohwy, J. (2013). *The predictive mind*. OUP Oxford.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815-836.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127-138.

Klein, S. B. (2013). The temporal orientation of memory: It's time for a change of direction. *Journal of Applied Research in Memory and Cognition*, *2*(4), 222-234.

Lewis, David K. (1973). Counterfactuals. Cambridge, MA, USA: Blackwell.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79-87.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 773-786.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans?. *Behavioral and brain sciences*, *30*(3), 299-313.

Werning, M. (2020). Predicting the past from minimal traces: Episodic memory and its distinction from imagination and preservation. *Review of philosophy and psychology*, *11*, 301-333.

Reimer, Ludmila (Ruhr-University Bochum) and Markus Werning (Ruhr University Bochum)

Co-Speech Gestures' Contribution to Sentence Meaning Composition

While traditional theories of semantics assume that the meaning of a sentence is compositionally determined by the meanings of individual words and how they are syntactically combined, dynamic views consider the evolution of meaning as an online process unfolding in time. While listening to an utterance word by word, the listeners generate predictions about the communicative intentions that speakers are about to express. With every word, these predictions might change. Whereas, according to traditional compositional semantics the meaning of a word is whatever it contributes to the meaning of a discourse, in this dynamic picture the meaning can be regarded as the change it makes to listeners' predictions of speakers' communicative intentions. Based on this dynamic picture, we designed two Bayesian Pragmatics models and three experiments to investigate how co-speech gestures contribute to sentence meaning composition.

We focussed on co-speech gestures co-occurring with sentences describing actions involving tools. Our stimulus material was constructed in a systematic way across experiments by creating sentences of the following structures:

General context: "The agent is v_{gen} -ing object *y*"

Specific context: "The agent is v_{spec} -ing object *y*."

Noun target: "To do so, they (x) are using ."

The *Noun* target sentences were constructed pairwise, featuring two instruments denoted by nouns that were both coherent continuations of the General context sentences and of their respective Specific contexts featuring the corresponding (congruent) verb v_{spe} of the noun E.g., the main telic component of a Förmchen ('cookie cutter') is ausstechen ('to cut out') and for Pinsel ('brush') it is bestreichen ('to glaze'). Both (cookie cutter and brush) are congruent continuations of the sentence:

Das Kind ist dabei, die Kekse zu backen. ('The child is baking cookies.')

But obviously not for each other's Specific context sentences since cookie cutters are usually not used for glazing, and brushes are not used for cutting out cookies:

Das Kind ist dabei, die Kekse auszustechen. ('The child is cutting out the cookies.')

Das Kind ist dabei, die Kekse zu bestreichen. ('The child is glazing the cookies.')

Next, we recorded a speaker, naïve to the experiments' purpose, reading the General context aloud and performing a gesture corresponding to v_{spec} and noun *n*. These videos were cut with uniform on- and offsets before and after the sentences, the speaker's face was blurred in post-processing.

As for our models, we needed the following values:

Tel (v, n) indicates whether the instrument denoted by the noun has the generic affordance denoted by the specific verb *v*, i.e., whether the telic component of the noun *n* is described by the specific verb *v*.

GloVe (n, v), measures the semantic similarity between *n* a noun *v* and a verb *v*, based on corpus linguistic co-occurrence statistics.

Recog (g, v), rated in Experiment 1, reflects how well a verb *v* describes a (silent) gesture and thus how well the gesture *g* is recognized as the action denoted by the verb *v*.

RatingText (n, v), rated in Experiment 2, measures the likelihood of the noun *n* in the written target sentence following the context sentence (either General or Specific) containing the verb *v*.

NRatingVideo (n, g (v), rated in Experiment 3, measures the likelihood of the written sentences containing the noun in the target sentence following a video containing the gesture *g(v)*, expressing the action denoted by the verb *v*, where the gesture accompanies the auditorily presented General Sentence.

Both our models are based on the following iteration of Bayes' rule (compare Werning et al. (2019), Werning and Cosentino (2017)):

(4) $P_V(n|g) = k \cdot P_V(g|n) \cdot P_V(n)$

After inserting functions of the values introduced above and subsequent logarithmization and linear approximation of the logarithmized functions, we get two models that differ in their identification of $P_V(n)$ *(NRatingText (n, v)* or *GloVe(n|v)*):

Model RTN:

$$(5)\ NRatingVideo\big(n, g(v)\big) = \begin{array}{l} z + a \cdot Recog(g(v), v) \\ +b \cdot Tel(v, n) \\ +c \cdot NRatingText(n, v) \end{array}$$

Model RTG:

$$(6)\ NRatingVideo\big(n, g(v)\big) = \begin{array}{l} z + a \cdot Recog(g, v) \\ +b \cdot Tel(v, n) \\ +c \cdot GloVe(n|v) \end{array}$$

In our three experiments, all differences of the reported scores were highly significant (p<.0001).

Experiment 1: We stripped the audio from our source videos resulting silent gestures being shown. Participants (N=40) had to rate on a 7-point scale whether the displayed verb described the gestures well. We paired the videos with the congruent and incongruent verbs v_{spec} and, as filler trials, with three randomly chosen verbs. Congruent target verbs scored 5,22 points (SD=0,60), incongruent targets scored 2,38 points (SD=0,61), and fillers scored 1,90 points (SD=0,57).

Experiment 2: No videos were used, only the written sentences constructed for our stimulus material. The participants (N=30) had to rate on a 7-point scale whether a sentence (Target Noun) was likely to follow after a given first sentence (either General context, Neutral condition, or Specific context). Congruent target verbs scored a mean rating of 6,48 points (SD= 0,50, incongruent targets 2,82 points (SD= 1,35), and neutral targets scored 5,36 points (SD= 1,16). This shows, that based on the linguistic input alone, the General context indeed allows both verbs v_{spec} as likely continuations but the Specific contexts do not.

Experiment 3. The participants (N=30) had to rate on a 7-point scale how likely they thought a displayed sentence (Target Noun) would follow the video. They were shown all three video variations of the source material, combined with the two corresponding Noun targets. Congruent target verbs scored a mean rating of 5,99 points (SD= 0,65), incongruent targets 2,98 points (SD= 0,92), and neutral targets scored 4,56 points (SD= 0,58).

Table 1: *Model comparison between Model RTN and Model RTG*

Model	N	df	RSE	R^2	R^2_{adj}	BIC	ΔBIC	AIC	AICc
Model RTN	188	184	0,6776	0,8449	0,8424	409,31	0	393,12	393,45
Model RTG	188	184	0,779	0,795	0,7917	461,74	52,44	445,56	445,89

Model RTN			**Model RTG**		
	p	β		p	β
$\boldsymbol{NRatingText(n, v)}$	$1{,}86e^{-13}$	0,3869272	$GloVe(n\|v)$	0,0855	0,7833378
$\boldsymbol{Tel(v, n)}$	$3{,}78e^{-13}$	1,5952457	$Tel(v, n)$	$< 2e^{-16}$	2,9167161
$\boldsymbol{Recog(g(v), v)}$	0,00175	0,1275582	$Recog(g(v), v)$	0,0132	0,1174883

As seen in Table 1, Model RTN offers the best fit for our data. We calculated the z statistic (Clogg et al., 1995)researchers should ask whether the coefficients associated with a given set of predictors change in a significant way when other predictors or covariates are added as controls. Simple calculations based on quantities provided by routines for regression analysis can be used to obtain the standard errors and other statistics that are required. Results are also given for the class of generalized linear models (e.g., logistic regression, log-linear models, etc. between the models to see whether their differences are significant, and found them to be highly significant (p<.0001). To conclude, we found that an iconic co-speech gesture makes a difference for a listener's probabilistic prediction regarding an upcoming instrument noun, and thus has a semantic effect on linguistic comprehension. The Model RTN that includes as a predictor how well the listener understands the action (denoted by the verb) as afforded by the instrument (denoted by the noun) predicts our data better than model RTG that instead includes as a predictor the co-occurrence statistics regarding the verb and the noun (GloVe).

References

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical Methods for Comparing Regression Coefficients Between Models. *American Journal of Sociology*, 100(5), 1261–1293.

Werning, M., & Cosentino, E. (2017). The interaction of Bayesian pragmatics and lexical semantics in linguistic interpretation: Using event-related potentials to investigate hearers' probabilistic predictions. In G. Gunzelmann,

A. Howes, T. Tenbrink, & E. Davelaar (Hrsg.), Proceedings of the 39th Annual Conference of the Cognitive Science Society (S. 3504–3509). Cognitive Science Society.
Werning, M., Unterhuber, M., & Wiedemann, G. (2019). Bayesian Pragmatics Provides the Best Quantitative Model of Context Effects on Word Meaning in EEG and Cloze Data. In A. Goel, C. Seifert, & C. Freska (Hrsg.), Proceedings of the 41th Annual Conference of the Cognitive Science Society (S. 3085–3091). Cognitive Science Society.

Reuter, Kevin (University of Zurich) and Lucien Baumgartner (University of Zurich)

Conspiracy theories are not theories

String theory, music theory, and critical race theory certainly differ in terms of their scientific, epistemic and socio-cultural standing. However, being all theories, they also share important properties: They are (i) investigated scientifically, (ii) learned and taught, and (iii) subject of approval and disapproval. In recent years, philosophers have begun to debate whether conspiracy theories are indeed theories or merely (systems of) beliefs (Dentith, 2019, Duetz, 2022, 2023, Napolitano, 2021, see also Napolitano & Reuter, 2021). In this talk, we present two corpus-linguistic studies with the aim to

a) identify markers of scientific theories that will allow us delineate "real" theories from theories in name only,

b) present corpus-based evidence demonstrating that conspiracy theories are mostly not considered theories,

c) highlight a significant overlap in the way both conspiracy theories on the one hand, and falsehoods and rumors on the other hand, are considered to be spreading misinformation.

Before attempting to determine whether conspiracy theories are truly theories, it is imperative to first address the question of what constitutes a theory. Regrettably, there is limited agreement regarding this matter (see for instance Abend (2008) and Duetz (2023)). Natural scientists rarely provide a definition of a theory, but instead emphasize the unique epistemic status of theories (National Academy of Sciences 1998, see also Popper 1963). They assert that theories undergo testing, confirmation, falsification, substantiation, refinement, and revision in response to the observations and experiments of the phenomena that the theories are meant to explain. In this paper, we adopt the latter practice-driven approach to examine the status of conspiracy theories. If it is found that people similarly test, confirm, substantiate, refine, and revise conspiracy theories, this would provide substantial evidence in support of the argument that conspiracy theories are indeed theories.

We examine whether conspiracy theories are regarded as theories through a linguistic analysis. If people engage in activities such as testing, confirming, and refining (conspiracy) theories, we can reasonably anticipate that they would also discuss these activities in talking and writing. Hence, examining the language used to describe the handling of theories and conspiracy theories can offer insight into their epistemic and scientific standing. One way to examine how people talk about what they do with theories and conspiracy theories is to collect a large number of phrases of the form "VERB [target term]", e.g., "test theories"', "share conspiracy theories", etc., from a corpus of choice.

Importantly, we need to compare the verbs preceding a whole range of different theories, in order to paint a fairly accurate and representative picture. This is what we have done in Study 1. We collected roughly 13,000 comments from the social media website Reddit, featuring phrases of the form "VERB [theory]", e.g., "studying game theory", "using music theory", and analyzed the relation between six control theories – critical race theory, fan theories, game theory, music theory, string theory, and theories (simpliciter) – and our target class conspiracy theories. We then examined and classified all those verbs standing before these theories. Many of these verbs fall into the following categories:

(a) Scientific verbs: confirm, create, develop, elaborate, falsify, prove, etc.

(b) Educational verbs: describe, discuss, explain, grasp, learn, study, teach, etc.

(c) Attitudinal verbs: accept, believe enjoy, hate, like, love, oppose, etc.

While all six control theories frequently featured verbs from the scientific and/or educational categories, verbs standing in front of "conspiracy theories" had a very low share of verbs belonging to these categories (see Figure 1).

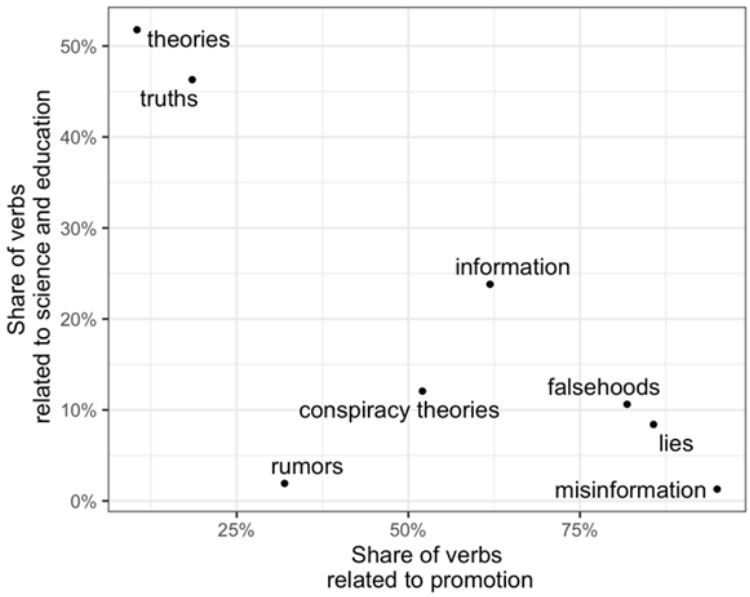


Figure 1: *Relation between different terms on a 2-dimensional space spawned by the components scientific and educational.*

Instead, verbs preceding "conspiracy theories" were dominated by a fourth category, which we call "spreading words". Instead of being elaborated, taught, studied, and applied, ordinary language users predominantly say that conspiracy theories are spread, pushed, and promoted.

Why would people so frequently talk about conspiracy theories in that way? An answer might be provided by looking at other phenomena that are spread, pushed and promoted, like falsehoods, misinformation and rumors. Thus, in order to further inquire into the similarities and dissimilarities of conspiracy theories on the one hand, and falsehoods and misinformation on the other, we decided to run a second corpus analysis in which we compare the categories of verbs preceding "conspiracy theories" with verbs occurring before terms such as "falsehoods". In Table 1, we list the 10 most frequent verbs for "conspiracy theories" and "falsehoods" from the NOW corpus.

Conspiracy Theories		Falsehoods	
Term	Number	Term	Number
promote	885	spread	748
spread	611	perpetuate	370
push	490	peddle	304
believe	441	contain	142
peddle	389	promote	136
embrace	243	publish	121
share	151	propagate	116
amplify	141	tell	113
debunk	136	use	100
espouse	135	debunk	92

Table 1: *A list of the 10 most frequent verbs in front of 'conspiracy theories', and 'falsehoods' on the NOW corpus.*

While a direct comparison between the ten most frequent verbs before "conspiracy theories" and "falsehoods" displays remarkable similarities, a more comprehensive comparative analysis is warranted once more. We therefore conducted a second corpus analysis using Reddit comments. The Reddit data for Study 2 consisted of roughly 12,000 new comments, in addition to the observations for "conspiracy theories" and "theories" previously used in Study 1. What we find is that "theories" and "truths" behave similarly, whereas "conspiracy theories" is much closer to falsehoods and rumors. Figure 2 compares the proportions of verbs related to science and education with those related to spreading. Here, again, we see "theories" and "truths" are used very differently from all other target expressions. Their share of spreading verbs is a lot lower, and their share of scientific and educational verbs is a lot higher compared to the majority of the other terms.

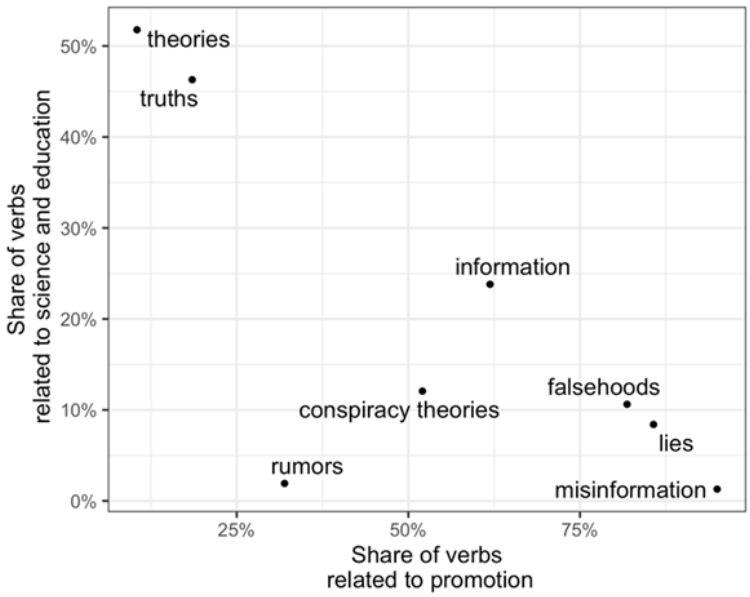


Figure 2: *Relation between the eight different terms*

The conclusion that conspiracies are not theories but rather like falsehoods or misinformation is likely to be met with resistance. In the last part of this talk, we consider various objections to our conclusion and discuss how those objections can be met.

References:

Abend, G. (2008). The meaning of 'theory'. Sociological theory, 26(2), 173-199.
Dentith, M R. X. (2019). Conspiracy theories on the basis of the evidence. Synthese, 196(6), 2243–2261.
Duetz, J. C. M. (2022). Conspiracy Theories are Not Beliefs. Erkenntnis, 1-15.
Duetz, J. C. M. (2023). What Does it Mean for a Conspiracy Theory to be a 'Theory'?. Social Epistemology.
Napolitano, M. G. (2021). Conspiracy theories and evidential self-insulation. In M. G. Napolitano (Ed.), The epistemology of fake news (pp. 82–106). Oxford University Press.
Napolitano, M. G., & Reuter, K. (2021). What is a Conspiracy Theory?. Erkenntnis, 1-28.
National Academy of Sciences (1998). Teaching about evolution and the nature of science. Joseph Henry Press.
Popper, K. (2014). Conjectures and refutations: The growth of scientific knowledge. routledge.

Ribeiro, Max Minden (Lund University)

Perceptual presence and the limit of metaphysical explanation

Look at something nearby. A book, a shadow, a friend. To have a visual perceptual experience is to experience that object as present to you. It is just there, materially real, occupying roughly the same region of space and time that you are. Compare this experience to one in which you close your eyes and visually imagine the same book or reenact in memory your perceptual experience of it. In these cases, the book, its colour, its shape, its location on the desk might all be brought before the mind, perhaps even in an imagistic equivalent to your original perceptual visual field. But something of the initial experience is missing. For in perceptual experience alone does it seem that the object is revealed to actually be there; a mind-independent constituent of present reality that experience grants access to. I term this phenomenal property of perceptual experience perceptual presence.

This paper has two objectives. The first is to offer a substantive definition of perceptual presence. This property is often characterised negatively, by pointing to contrasts with other kinds of experience. I provide a positive definition that serves as criterion for the property's instantiation:

Perceptual presence is instantiated if and only if:

(1) The object is presented as a part of material mind-independent reality and

(2) the experience seems, by the very kind of experience it is, to reveal how things are.

The second objective is to assess competing metaphysical explanations of perceptual presence. The two most prominent accounts of perceptual experience are naive realism and representationalism. Proponents of both positions have argued that their view is especially well placed to explain or accommodate presence. I contend that there are limits to how informative such a metaphysical explanation can be. Nonetheless, both a naive realist and certain representationalists – those that explains presence in terms of representational attitude – will have equal claim to explaining presence. By contrast, a representationalist that appeals to representational content faces difficulties. The latter is the kind of position espoused by Boyd Millar (2014a, 2014b).

This paper has three sections. In Section 1, I motivate my definition of perceptual presence and defend it against counterexamples. In particular, I show that visual imagination, episodic recollection, the experience of watching live television, pictorial experiences, pain experiences and insight experiences instantiate at most only one of the two conditions.

In Section 2, I ask whether naive realism can accommodate perceptual presence. After presenting a naive realist account of presence I introduce two lines of criticism. First, I develop a sceptical concern from O'Conaill (2017), namely that the naive realist explanation is unsatisfactory in that it fails to show why the phenomenology of presence necessarily follows from their metaphysical picture. For why should it be that an object-dependent account invokes a phenomenology of presence? I respond that it is not object-dependence but the nature of the acquaintance relation that does the explanatory work. I agree with Millar (2014a) that there is a certain circularity in the way the naive realist specifies and appeals to this relation. But I propose that the circularity is benign and indeed desirable if the naive realist is to avoid the kind of explanatory gap that O'Conaill's criticism points to. Second, I consider Millar's (2014a) objection that a naive realist explanation of presence (or 'phenomenological directness', as he terms it) will struggle to accommodate perceptual objects' seeming mind-independence. I show that Millar's argument depends on his conceiving of the phenomenology of presence as conjunctive, which I argue is a mistake.

In Section 3, I turn to representationalism. I introduce two strategies the representationalist might adopt to explain presence, namely in terms of representational content and representational attitude. Millar's direct causal content view (2014a, 2014b) illustrates the former, while Mohan Matthen's account of the 'feeling of presence' (2005, 2010, 2014) provides a useful example of the latter. Millar intends his position to provide a non-circular explanation of presence. I argue that not only does this account fall short in accommodating object-immediacy and object-distinctness, (the two phenomenal properties the conjunction of which realise perceptual presence according to Millar), the non-circularity it aspires toward leaves it unsatisfying as an explanation of phenomenology. By contrast, attitude-explanations (such as Matthen's) provide an equivalent explanation to the naive realist one that I develop in Section 2. Both have limited explanatory resources, but in terms of accommodating presence, there is nothing to choose between these two positions.

Roessler, Johannes (Department of Philosophy, University of Warwick)

Perceptual knowledge and doxastic responsibility

Recent work on the nature of our responsibility for what we believe is marked by a broad consensus on a basic question and a variety of opinions on how to develop the consensus view. It is widely agreed that since we can appropriately be held answerable for what we believe, there must be a sense in which we are 'active' in relation to our current beliefs. Believing something, according to the consensus view, involves an exercise of 'rational agency or 'self-determination. I will call this the Activity thesis (AT). A multitude of different suggestions have been aired about how to understand the operative sense of "activity" or "self-determination". For some, the key to AT lies in the relation between belief and active deliberation. (Moran 2001) Others focus on what they see as the central role of acts of judging in the formation of beliefs, with different accounts having been proposed of the nature of such acts. (McDowell 1994, 1998a; Hughes 2013) Yet others suggest

that there is a sense in which believing itself — a state, not an event or process — amounts to a rational activity. (Boyle 2009a, 2011)

In this talk I present a challenge to AT, by connecting the question of the sense (if any) in which believing may be said to be 'active' with a topic that has been conspicuous by its absence from recent discussions of that question: the relationship between belief and propositional knowledge. Two major contributions to the two areas — Timothy Williamson's Knowledge and its Limits and Richard Moran's Authority and Estrangement — were published within a year of each other, but despite the temporal proximity, there has been little interaction between the debates generated by these works. One might find this surprising, simply insofar as both debates are concerned with the nature of belief. Still, it is a good question whether there are any substantive connections between the two sorts of issues. Might they simply be orthogonal to each other? I want to suggest that they are not. My talk will be in two parts. In the first part I set out a 'primitivist' view of perceptual knowledge, from which my challenge to AT develops. In the second part I articulate the challenge and canvass some responses to it.

Primitivism

Current work on AT, so I will argue, is informed by a dubious conception of the relation between two kinds of questions: 'Why do you believe that p?' and 'How do you know that p?' The unspoken assumption in this work is that the former question is always explanatorily prior: our understanding of how someone knows that p invariably turns on the explanation of why they believe that p. I will suggest that this 'belief first' approach falsifies the relationship between perception and knowledge, as conceived in our ordinary practice of probing claims to knowledge. As ordinarily conceived, the explanatory connection between perception and knowledge is basic or primitive: it is not to be analyzed by reference to some underlying explanatory link between perception and belief (whether theorized along internalist or externalist lines). For example, our ordinary understanding of how you know that a certain animal is a pig (when the pig is right there in front of you) makes no reference to either your reason for belief or a non-rational cause of your belief. It makes no reference to your belief at all. Rather, the explanation is provided by your exercising capacities for visual knowledge, such as the capacity to tell a pig when you see one. (See Austin 1961, 1962; Millar 2019; Stroud 2011) No doubt in coming to see and so know that the animal is a pig you also come to believe that it's a pig. This is because knowledge entails belief. But the entailment is compatible with the explanatory priority of knowledge. A good explanation of why you believe it's a pig is that, seeing the pig and exercising your visual-recognitional capacity for pigs, you come to know it's a pig.

There will be no time to attempt a detailed defence of the primitivist analysis, but I will try to say enough to bring out some of its attractions and to convince the audience that the analysis is worth taking seriously. My main concern, though, will be primitivism's bearing on AT.

The challenge to at

Here is an initial reason to think that a primitivist analysis of our ordinary conception of perceptual knowledge spells trouble for AT. Advocates of AT assume that our practice of holding each other answerable for our beliefs is inextricably connected with the demand for reasons. In a nutshell: it is assumed that our practices of doxastic answerability show beliefs to be under the sway of our capacity to assess the force of reasons, and thus to exercise a form of doxastic control or self-determination. Yet, if the explanatory connection between perception and knowledge is primitive, we will be able to produce credentials for a perceptual belief without invoking any reason for which we hold the belief. We could instead say things like 'I can just see that there is a pig in front of me' or 'I can tell a pig when I see one', where these statements should be taken to gesture towards an account of how seeing a pig, in concert with our visual-epistemic capacities, gives us knowledge that there is a pig in front of us. On the face of it, the knowledge-producing role of perception dispenses us from the task of exercising judgement as to the probative value of our evidence.

I will try to spell out the challenge in more detail, and then briefly discuss some lines of response. My conclusion will be that while AT may be modified in a way that would make it compatible with primitivism, the modified version would not be able to do the work commonly assigned to AT.

Rostworowski, Wojciech (University of Warsaw), Katarzyna Kuś (University of Warsaw) and Bartosz Maćkiewicz (University of Warsaw)

Non-Doxastic Attitudes, Assessment, and Conjunction

In the talk, we argue that the context of ascribing attitudes to an agent can strengthen the interpretation of typical logical connectives. We focus on the connective 'and'. Our argument is supported by empirical data obtained by experimental methods.

Reports of attitudes – i.e., ascriptions such as 'S believes that p', 'S is glad that p', 'S wants p', etc. – pose many difficulties to semantic theory. They are sensitive to various modifications of the embedded clause that do not affect the truth-conditional content of the clause itself, yet affect the truth-conditional content of the whole ascription. These problems include, for instance, non-replacability of co-referring names (see Frege 1892, Heck 2012) and failing to support logical entailments (Cresswell & Bäuerle 1989). Furthermore, the verbs expressing non-doxastic attitudes (like desires, fears, etc.) exhibits a complex filtering of the presuppositions of the embedded clause (cf. Heim 1992, Maier 2015), which is arguably an aspect of a more general problem that the truth conditions of such ascriptions are sensitive to the information structure of the embedded clause (see Rostworowski 2018 and Rostworowski, Kuś, Maćkiewicz 2023). In our talk, we identify another problem, which concerns the behavior of logical connectives under non-doxastic attitude verbs. We focus on conjunctions expressed by 'and'. According to the classical view, 'and' is semantically equivalent to the logical connective '∧', which entails that 'p and q' is true exactly if both p, and q are true. (Occasionally, speakers may interpret 'and' as expressing a stronger connection, i.e., temporal or causal; for some empirical research see Noveck et al. 2008, Blochowiak and Castelain 2018). Yet, the conjunction behaves differently when embedded under a non-doxastic attitude verb. On the one hand, a sentence the form, e.g.,

(1) S is glad that p and q,

has a distributive reading, i.e., people tend to infer from it that S is glad that p and likewise S is glad that q. On the other hand, in an appropriate context, we may accept a report of the form in (1) to a certain degree, if S is glad that q but not particularly happy that p and only "accepts" p because it enabled, for instance, to accomplish q. For example, Anne may not be happy that she went to the forest during the rain and got wet, but she may eventually be satisfied because she found a lot of mushrooms, which was her goal. So we are willing to accept (2) below, all things considered:

(2) Anne is glad that she had a walk in the rain and found a lot of mushrooms.

Such an evaluation of the ascription is semantically complex as it does not (only) include the subject's attitude towards both facts, but (also) an assessment which fact is actually more important to the agent, which determines the conclusion that Anne has an overall positive attitude to the content expressed by the embedded conjunction.

In the talk, we present a series of experiments exploring the indicated phenomenon. Our main experiment confirms two predictions: I) people evaluate an attitude ascription with an embedded conjunction differently depending on whether the attitude is doxastic or non-doxastic in nature; II) in the cases where one of the embedded conjuncts is false, people are more likely to ascribe a non-doxastic attitude to the agent than the doxastic attitude. The experiment had a within-subject design with a type of conjunctive sentence as a factor that has four levels: two types non-doxastic attitudes (expressed by the verbs 'want' and 'glad'), doxastic attitude ('know') and an nonembedded conjunction. The study had the form of a questionnaire in which participants (N = 100) were asked to read short fictional stories and performed two true-value-judgment tasks (evaluation of the target sentence plus evaluation of the first conjunct which was false in light of a scenario) followed by confidence rating. In the non-doxastic attitudes, about ¼ of the participants indicated that the attribution is 'true' ('glad': 25.64%, 'want': 28.57%) whereas for doxastic attitudes and nonembedded conjunctions almost no subjects judged them as 'true' ('know': 6.02% and: 3.75%, $\chi 2(3) = 29.93$, $p < 0.001$). This statistically significant result supports both predictions I) and II).

The findings of our experiments suggest that people do not fully reject attitude ascriptions such as (2), provided that only one of the embedded conjunct expresses a proper object of the agent's attitude. It is hard to explain this result by appealing only to pragmatics, in particular, to charity which generally postulates not to reject any statement that contains a 'partial truth'. If that were the case, people should likewise weakly accept the doxastic attitude ascriptions and nonembedded conjunctions in which one conjunct

was always false and the other true, according to the presented story. The result, in turn, is compatible with our hypothesis that the evaluation of a non-doxastic attitude ascription includes an assessment of the importance of the embedded conjuncts from the agent's viewpoint, allowing for natural variations in such assessments between subjects. This furthermore suggests that the truth of a non-doxastic attitude ascription 'S ATT that p and q' does not straightforwardly depend on the truth of the ascriptions 'S ATT that p ' and 'S ATT that q'.

References:

Blochowiak, J., & Castelain, T. (2018). 'How logical is natural language conjunction? An experimental investigation of the French conjunction *et*', in *Language, evolution and mind: Essays in honour of Anne Reboul* (ed. Pierre Saint-Germier), College Publications, 97–125.

Cresswell, M., Bäuerle, R. (1989). 'Propositional attitudes', in *Handbook of Philosophical Logic* (ed. F. Guenthner and D.M. Gabbay) Vol IV, Dordrecht, Reidel, 491–512.

Frege, G., 1892. 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kriti*k, 100: 25–50.

Heck, R. (2012). 'Solving Frege's puzzle', *Journal of Philosophy*, 109: 132-174.

Heim, I. (1992). 'Presupposition projection and semantics of attitude verbs', *Journal of Semantics* 9(3): 183–221.

Maier, E. (2015). 'Parasitic attitudes', *Linguistics and Philosophy* 38, 205-236.

Noveck, I. A., & Reboul, A. (2008). 'Experimental pragmatics: A Gricean turn in the study of language', *Trends in Cognitive Sciences* 12(11), 425–431.

Rostworowski, W. (2018). 'Descriptions and non-doxastic attitude ascriptions', *Philosophical Studie*s 175: 1311–1331.

Rostworowski, W., Kuś, K., & Maćkiewicz, B. (2023). 'Semantics of Non-Doxastic Attitude Ascriptions from Experimental Perspective', *Proceedings in Experiments in Linguistic Meaning* 2 (eds. T. Knowlton, F. Schwarz, A. Papafragou), 241–251.

Rowson, Rebecca (Centre for Philosophical Psychology, University of Antwerp)

The Object of Surprise

If we were to arrive home and find several of our friends jumping out from behind the furniture, it is likely we would feel surprised. It is natural to treat experiences like these as emotional experiences. As with other emotions, surprise has certain action tendencies and physiological changes associated with it, it is associated with particular expressions, it stands in relation to an intentional object (we are surprised by something), and it has a particular phenomenology.

Despite this, philosophers of emotion have been relatively quiet on the subject. Within philosophy, the discussion of surprise has been mostly confined to epistemology. Epistemologists argue that for something to surprise us, it must be unexpected i.e. we must assign it a low prior probability (Horwich, 1982).

In this paper, I argue that if we are to treat surprise as we do other emotions in philosophy, we must be sensitive to the distinction between the cause and object of surprise. Existing philosophical accounts of surprise have conflated these two things. But I argue that there are three significant problems for thinking about the object of surprise in terms of prior expectations. I end by introducing two alternative candidates for the object of surprise based on an extended body of work on surprise in psychology. Contrast and sense-making theories stand beside expectation theories in psychology but have yet to be appreciated in philosophical discussion.

I start the paper by outlining the longstanding distinction between the cause of an emotion and the object of an emotion (Hume, 1739; Kenny, 1963). Emotions, like other mental phenomena, have causes. Being kicked is the cause of the pain one feels, a lie can be the cause of the hurt, etc. Causes of emotions are often listed as reasons why a particular emotion occurred. Emotions are also said to have intentional objects – things that they are directed towards. A person is afraid of the tiger and sad about the lie. We can further distinguish an emotion into its formal and particular object. Not only are emotions directed towards things like the tiger (the particular object), they also involve evaluations of those things, such as the evaluation that the tiger is dangerous (the formal object) (Teroni, 2007, 396).

It is with respect to the object of emotion that we determine whether an emotion is fitting or unfitting. An emotion is fitting if the evaluation we make of some (particular) object matches the way things are in the

world (Tappolet, 2016). My fear of the tiger is fitting if the tiger is in fact dangerous. What makes an emotional response fitting is often what's discussed by philosophers trying to distinguish one emotion from another.

In a recent paper, Baras and Na'aman propose to do this for surprise. They argue that surprise is fitting if it is unexpected and significant to the agent (Baras & Na'aman, 2021, 207). For the most part I will not be concerned with the significance condition here, but rather with the expectation condition. As I indicate above, the expectation condition has a long tradition in epistemology (Good, 1984; Harker, 2012; Horwich, 1982; Manson & Thrush, 2003; Olsson, 2002; Shogenji, 2021; White, 2000). However, by putting their account in terms of fittingnesss, and given that something is fitting or unfitting in relation to its object, Baras and Na'aman implicitly treat expectations as part of the object of surprise and not merely its cause.

I next present three problems for this account, and any which treats expectations as the object of surprise. Firstly, I argue that treating the object of surprise in terms of prior expectations renders surprise either self-defeating or endlessly fitting. We can cash out the account by saying that surprise is fitting if some event (say, a surprise party) is unexpected. But once the unexpected event occurs, it is no longer unexpected. As such, surprise cannot fittingly persist at all. This has the implausible consequence that, once occurring, surprise is never fitting. The proponent of the expectation-as-object account might adjust the theory by saying that surprise is fitting if the surprise party was unexpected. The conditions that make surprise a fitting emotion obtain in the past rather than the present and we can avoid the previous worry. However, once the surprising event has occurred, what makes surprise fitting will remain fixed – it will always have been unexpected. This distinguishes surprise from other emotions. While emotions like remorse may have particular objects that obtain in the past (say, a lie that was told), their formal objects (say, the evaluation that the lie is harmful) remain flexible. With other emotions, there is always scope for the emotion to become either fitting or unfitting.

Secondly, I argue that if the object of surprise were prior expectations, we would not be able to feel surprised in the wake of transformative experiences. If, when we are surprised, we turn our attention towards a prior probability rating, this restricts what we can be surprised by to things for which a prior probability rating is possible. Insofar as transformative experiences change our preferences such that we cannot assign accurate probabilities to our experiential states before or after the experience (Paul, 2014), any surprise we feel after such experiences cannot have as its object something that transverses the transformation.

Finally, I argue that results from empirical research on the degree to which we feel surprised by different events cannot be adequately captured by either an expectation condition or a significance condition. While results from experiments by Teigen & Keren (2003) cannot be accounted for by expectations, they suggest that explaining surprise in terms of contrast better accounts for the empirical evidence. I look at this account as well as proposals which treat surprise as a kind of 'sense-making' task (Foster & Keane, 2013; Maguire et al., 2011). I suggest that these accounts found in psychology provide us with insights for how philosophers of emotion might recapture the object of surprise and avoid the three problems I have raised.

References

Baras, D., & Na'aman, O. (2021). What Makes Something Surprising? *Philosophy and Phenomenological Research*, phpr.12820. https://doi.org/10.1111/phpr.12820

Deonna, J. A., & Teroni, F. (2012). *The Emotions: A Philosophical Introduction*. Routledge.

Foster, M., & Keane, M. T. (2013). Surprise: Youve got some explaining to do. ArXiv:1308.2236 [Cs]. http://arxiv.org/abs/1308.2236

Good, I. J. (1984). A Bayesian Approach in the Philosophy of Inference. *British Journal for the Philosophy of Science*, 35(2), 161–166. https://doi.org/10.1093/bjps/35.2.161

Harker, D. (2012). A surprise for Horwich (and some advocates of the fine-tuning argument (which does not include Horwich (as far as I know))). *Philosophical Studies*, 161(2), 247–261. https://doi.org/10.1007/s11098-011-9732-3

Horwich, P. (1982). *Probability and Evidence*. Cambridge University Press.

Hume, D. (2007). *A Treatise of Human Nature: A Critical Edition*. Oxford University Press.

Kenny, A. (1963). Action, Emotion And Will. Ny: Humanities Press.

Maguire, R., Maguire, P., & Keane, M. T. (2011). Making sense of surprise: An investigation of the factors influencing surprise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 176–186. https://doi.org/10.1037/a0021609

Manson, N. A., & Thrush, M. J. (2003). Fine-Tuning, Multiple Universes, and the 'This Universe' Objection. *Pacific Philosophical Quarterly*, 84(1), 67–83. https://doi.org/10.1111/1468-0114.00163

Olsson, E. J. (2002). Corroborating Testimony, Probability and Surprise. *British Journal for the Philosophy of Science*, 53(2), 273–288. https://doi.org/10.1093/bjps/53.2.273

Paul, L. A. (2014). *Transformative Experience*. Oxford University Press.
Shogenji, T. (2021). Probability and proximity in surprise. *Synthese*, 198(11), 10939–10957. https://doi.org/10.1007/s11229-020-02761-6
Tappolet, C. (2016). *Emotions, Values, and Agency*. Oxford University Press. https://doi.org/10.1093 acprof:oso/9780199696512.001.0001
Teigen, K. H., & Keren, G. (2003). Surprises: Low probabilities or high contrasts? *Cognition*, 87(2), 55–71. https://doi.org/10.1016/s0010-0277(02)00201-9
Teroni, F. (2007). Emotions and Formal Objects. *Dialectica*, 61(3), 395–415.
White, R. (2000). Fine-Tuning and Multiple Universes. *Noûs*, 34(2), 260–276. https://doi.org/10.1111/0029-4624.00210

Scarafone, Antonio (LMU Munich)

Communication and Mindreading in Infancy

It is widely believed that human communication essentially consists in intending to make one's own intentions and beliefs recognised by the audience. Call this view 'Gricean'. It is often claimed by philosophers, linguists, and psychologists alike,1 that infants are Gricean communicators, and that this explains how they can communicate so effectively at the prelinguistic stage (see, e.g., Tomasello 2008, 2019), as well as the facility with which they acquire the meaning of words and constructions of a natural language (see, e.g., Bloom 2000, Tomasello 2003). On this view, explicit reasoning about intentions and beliefs is a prerequisite for learning how to speak. It is often argued (see, e.g., Harris 2019) that experiments on infant 'mindreading' have amply vindicated this Gricean stance from an experimental point of view. In this talk, I argue that experiments on infant mindreading, far from constituting evidence in favour of the Gricean view, in fact undermine it.

To make my case, I will focus on an important study conducted by David Buttelmann and colleagues (2009), and on its troubled history of replication attempts. In this experiment, in the presence of the infant a toy is moved from Box 1 to Box 2, while an adult is either present and witnessing the transfer (True Belief condition) or absent (False Belief condition). In the following phase, this same adult tries to open Box 1, either truly believing that the toy is not there, or falsely believing that the toy is still there. In the True Belief condition, infants tend to help the adult by opening Box 1; presumably, infants assume that the adult wants to open this box despite knowing that the toy is not there. In the False Belief condition, infants tend to help the adult by opening Box 2; presumably, infants assume that the adult wants to open Box 1 because the adult falsely believes that the toy is there, and so to help the adult they open the Box where the toy really is. These results suggest that infants take into account the adult's belief for individuating the adult's goal, and help accordingly.

This and other similar experiments, I argue, constitute the best candidate source of evidence in favour of the claim that infants reason explicitly about intentions and beliefs (Carruthers 2020), as required by the Gricean theorist. However, the validity of the experiment by Buttelmann et al (2009) has been convincingly, though not conclusively, questioned (Priewasser et al. 2018). Up to now, most replication attempts, be them direct or conceptual, have yielded either partial or null results (e.g., Poulin-Dubois et al. 2018). Assuming that a refined version of the experiment will be valid, directly replicating the original results will require great care, as acknowledged by the authors of the original experiment themselves (Baillargeon et al. 2018).

Indeed, the mindreading effect disappears, for instance, if infants happen to be too close to the boxes, or if they are familiar with the adult (e.g., Crivello & Poulin-Dubois 2018). Other experiments on infant mindreading, which use as dependent measures either looking time, anticipatory looking, or neural response, constitute an even more indirect source evidence in favour of the Gricean stance, and the mindreading effect they detect is similarly fragile. The fragility of the effect is its Achille's heel: if infant mindreading is so fragile, it cannot systematically support infant communicative success as the Gricean theorist assumes that it does. Indeed, if infant mindreading is equated with psychological reasoning of the Gricean variety, it follows that infants do not communicate in Grice's sense whenever they interact with their primary caregiver, or whenever they are close to the objects that they are communicating about. These conclusions are not acceptable for the Gricean theorist, and in fact constitute a reductio ad absurdum

of this stance, or so I will argue. In conclusion, I will point out that this mismatch between mindreading and communicative abilities is no reason for despair. Indeed, there are perfectly viablealternative conceptions of infant communication which do not postulate any robust capacity for psychological reasoning on the side of the child (reference anonymised).

Schidelko, Lydia Paulin (University of Göttingen, Germany), Marina Proft (University of Göttingen), Marlene Meyer (Department of Cognitive Developmental Psychology, Georg-August-Universität Göttingen (Germany)), Jan Engelmann (University of California, Berkeley) and Hannes Rakoczy (University of Göttingen)

Reasoning about Possibilities Under Epistemic and Physical Uncertainty

Influential recent accounts claim that children acquire concepts of epistemic and other kinds of modality relatively late, not until the age of four or older (Leahy & Carey, 2020; Redshaw & Suddendorf, 2020). The main empirical support for these claims comes from studies in which children are faced with an uncertain situation, for example, in which of three cups two stickers are hidden (Mody & Carey, 2016) or which of two objects is in a given box (Kloo et al., 2017). Across tasks, children consistently fail to appreciate their uncertainty about alternative possibilities or to choose a certain over an uncertain option. In light of contrasting results from possibility understanding in infancy (e.g., Cesana-Arlotti et al., 2018), two pressing research questions arise. First, do these results of preschooler's failure present false negatives? Second, what are the cognitive foundations of modal thought?

In five studies, we aim to investigate whether these results of late developing modal reasoning persist in adapted measures and optimal test settings and compare children's performance in modal reasoning tasks to their developing meta-representational abilities.

In Study 1, three- to five-year-old children (N= 90) were tested in their ability to hold modality representations in a simplified version of the forked tube task (physical uncertainty; Redshaw & Suddendorf, 2016) and a partial exposure task (epistemic uncertainty; Kloo et al., 2017). We compared children's performance in these modified modality measures with their developing first- and second-order false belief reasoning as a measure of meta-representational thinking (Wimmer & Perner, 1983). In Study 2 (N = 100, age: 3-6 years) and Study 3 (in data collection, final sample N = 90; age: 3-5 years) we test for children's possibility representations under epistemic and physical uncertainty in a minimal contrast paradigm. To this end, children are asked to prepare for an event that happened but they are ignorant about (epistemic uncertainty) or a future event (physical uncertainty) in a slides task (adapted from Beck et al., 2006; Robinson et al., 2006) and in a version of a forked tube task (Redshaw & Suddendorf, 2016). Again, children's performance in these tasks is compared with their developing first- and second-order false belief reasoning. In Study 4 (N = 66; 3- and 5-year-olds) and Study 5 (in data collection, final sample N = 60, age: 3 years) we administer versions of a partial exposure task (epistemic uncertainty; Rohwer et al., 2012; Kloo et al., 2017) and a three-cups task (epistemic uncertainty; Mody & Carey, 2016) in a social, cooperative test situation. A cooperative and communicative context, in which children have to give reasons and justifications towards others, might enhance young children's metacognitive insights and their modal reasoning abilities under epistemic uncertainty.

The results of these studies will be discussed with regard to (1) accounts claiming modal thought in infancy (e.g., Cesana-Arlotti et al., 2018), contrasting (2) accounts claiming a late development of modal thought (Leahy & Carey, 2020) and (3) accounts claiming meta-representational foundation of modal reasoning (e.g., Redshaw & Suddendorf, 2020).

Scholz, Sebastian (Heinrich-Heine-Universität Düsseldorf)

Conceptual Spaces: A Solution to Goodman's New Riddle of Induction?

The New Riddle of Induction is that we cannot easily distinguish "good" (viz. projectible) predicates that are allowed for inferences from peculiar ones that are prohibited – such as the infamous Goodman-predicate 'grue'. This is at least one common way of framing the problem. But the epistemological concern at stake is this: What is the justification of our inferential practice?

Since Goodman posed his puzzle in 1955, many solutions have been advanced. A particularly intriguing proposal comes from recent cognitive research. Inspired by Quine (1969), Peter Gärdenfors (1990, 2000, 2011, 2019) argues the good predicates are those referring to natural entities – and that naturalness can be captured in terms of similarity. In contrast to Quine, however, he does not primarily rely on logic, but geometry. His innovation is to represent similarity as a function of distance in so-called Conceptual Spaces. As similar objects form clusters in such spaces, Gärdenfors defines properties and concepts as regions – and regards them as natural if they display a degree of topological cohesion, e.g., if they possess the geometric property of convexity. A region is said to be convex iff for any points A and B within the region it holds there is no point C between A and B that is not also in the region.

But even assuming the convexity criterion can successfully discern good predicates from bad ones (which might be contested on independent grounds), I will argue it does not offer any substantial justification of our inferential practice – and so falls short as a solution to the riddle. My argument is based on Gärdenfors' views on the philosophy and metaphysics of science and takes the form of a dilemma: He can either take his brand of instrumentalism seriously, or tone it down by talking about the structure of external reality. If he chooses the first option, this leads into radical epistemic relativism – a consequence he claims to avoid. If he chooses the second option, then his convexity criterion becomes obsolete – at least without additional ontological presumptions. Whichever option is preferred, his account of naturalness is at best incomplete if it is meant to resolve the riddle.

I shall briefly spell out some of the details, starting with the characteristics of the approach on which my argument is built. As an adherent of Cognitive Semantics, Gärdenfors maintains that "The epistemological question of the relation between the conceptual structure and the world must be kept separated from the semantic question of the relation between linguistic expressions and the conceptual structure." (Gärdenfors 2000, sec. 5.1.3). He believes an exhaustive semantics of projectible predicates can be given with reference to conceptual structures (i.e. convex regions) only, while those structures do not represent the external world. This broadly internalist and neo-Kantian (see Gärdenfors 2000, footnote 161) perspective is coupled with sweeping instrumentalism: "[…] I avoid questions about how real the dimensions of conceptual spaces are but view them as instruments for predictive and constructive purposes […]. The important thing is that we can do things with [conceptual spaces]." (Gärdenfors 2000, sec. 1.10). All this amounts to a purely cognitive notion of naturalness, as opposed to the ontological naturalness that David Lewis, for instance, is concerned with: "Nor should it be said […] that as a contingent psychological fact we turn out to have states whose content involves some properties rather than others, and that is what makes it so that the former properties are more natural. (This would be a psychologistic theory of naturalness.)" (Lewis 1989, p. 377).

Now, Goodman's Riddle is about the projectibility of predicates from one set of cases to any other set. On Gärdenfors' account, both these sets are part of the internal makeup of cognitive agents. It would appear, therefore, that a justification of inferences cannot "reach out" into the world, but is only possible relative to said makeup. Indeed, this account comes very close to Goodman's own (non-)solution: "Instead of providing a theory that would ultimately justify our choice of predicates for induction, he develops a theory that provides an account of how we in fact choose predicates for induction and projection. […] [He] makes projectibility essentially a matter of what language we use and have used to describe and predict the behavior of our world." (Cohnitz & Rossberg 2022, sec. 5.4). The passage is about Goodman, but if we replace „language" with „concepts" we get an adequate characterization of what Gärdenfors proposes. The problem gets even clearer when one realizes that the dimensions of conceptual spaces are not just given by nature, but are established via interpretation – or stipulation – by scientists. If we take this seriously, then epistemic relativism appears to follow suit.

Gärdenfors is aware of this hazard and, against it, puts forward a pragmatic stance according to which

at least some fundamental conceptual spaces adapt to the environment, becoming 'viable' through evolution. In my view, this is insufficient, not least because there are many examples where evolution has systematically distorted our thinking. For a substantive justification of inference, we need to speak more directly about the structure of mind-independent reality – which brings me to the second horn of the dilemma. For if Gärdenfors were to follow this suggestion, he would run into a different problem: Embracing mind-independence raises the need for an ontological demarcation of natural entities. His convexity criterion, however, pertains to the internal make-up of cognitive agents and prima facie cannot be applied to the world. Without further presumptions (e.g. correspondence of cognitive distance and ontological similarity), it is of little use for present ends.

So convexity as a purely cognitive criterion results in relativism, conflicting with a solution that includes justification of our inferential practice. Yet even if mind-independent components are accepted, modifications of the Conceptual Spaces approach are needed to meet this goal. At any rate, in its current form, the approach does not provide a satisfactory solution to Goodman's Riddle.

Schröder, Jan-Luca (University of Osnabrück; Università degli Studi di Trento) and Annette Hohenberger (University of Osnabrück Institute of Cognitive Science)

Effect of Attention to fractal and non-fractal images in a novel two choice SART-oddball task

In the current study, we investigated a special class of hierarchical geometrical structures, so-called fractals, in terms of attention and perception. Studies regarding the perception of fractals are a topic of growing literature in cognitive and life sciences. Fractals seem to play a unique role, both in terms of processing and in nature. Recent studies have shown that humans seem to have a special capacity regarding the processing of such recursive structures (see Fitch (2014)).

Previously, however, recursive structures were primarily studied in terms of the linguistic domain, and it was even proposed that this phenomenon might be unique to this domain. However, recent studies have started to also provide evidence for recursive processing in the visual domain, on which we focus as well. Our study was inspired by the study of Martins et al. (2014), which looked at the neural activation patterns when viewing fractal and non-fractal images. However, they did not go into detail concerning behavioural measures of perception of fractals, which are the focus of our study.

We wanted to investigate the unique human capability of perceiving and, what is new, attending to fractal stimuli.

Participants were presented with either fractals (i.e., self-similar structures) or similarly build non-fractals (i.e., non-self-similar structures) during a novel-paradigm: a two-choice Sustained Attention to Performance (SART) oddball task, referred to as 'oddball task' in the following (see Figure 1 and 4). Additionally, we performed two other tasks, which were replications of the Visual Recursion Task (VRT) and Embedded Iteration Task (EIT) in Martins (2012) and Martins et al. (2014), in order to further study the effects of fractal processing in a generation-based task.

This study was conducted as a browser based online study with 97 healthy participants and took approximately 30 minutes per participant. Participants were first instructed about the concept of fractals, before the start of the actual trials, in order to counteract possible effects of different levels of knowledge about fractals. We looked at the behavioural results, recorded during the experiment, namely the reaction time (RT) and the correctness for each trial in the tasks. In the oddball task, the participants were presented with one image at a time and had to decide with a button press ('G' or 'H' key) to which of the two categories (fractal or non-fractal) the image belongs. In each condition, one of the categories was presented as the standard stimuli, while an image of the other category was presented after every one to eight images from the first category, respectively (see Figure 1).

We hypothesized an effect of fractals vs non-fractals on both RT and correctness, which would result in higher correctness and faster RT. We further hypothesized, regarding attentional effects, that the fractal

oddball stimuli would result in an effect on RT and correctness that this would be distinct from the general effect of the oddball stimuli (fractal and non-fractal). We expected this to result in higher correctness and faster RT, compared to the non-fractal oddball stimuli and the non-oddball stimuli, as the fractal oddball stimuli would 'grab the attention better'.

The analysis was performed using Bayesian statistics. Our results suggested that fractals appear to have a strong effect on both RT and correctness, resulting in overall shorter RT and higher correctness compared to similar non-fractal images (see Figure 2 and 3).

We also found similar strong effects in the EIT and VRT task, where fractal trials resulted in significantly higher correctness and faster RT. In addition, the oddball stimuli seem to have an effect on both the correctness and the RT compared to images from the same category as the standard. Interestingly, the strength and direction of this effect differs between fractals and non-fractals. Fractal oddball stimuli are more correct than non-fractal oddball stimuli, but result in slower RT. However, while the non-fractal oddball stimuli resulted in faster RT and higher correctness compared to the non-fractal standard stimuli, the fractal oddball stimuli resulted in slower RT and lower correctness, compared to the fractal standard stimuli. These results further indicate a special role of fractal structures in attention.

Our results provide strong evidence for the effect of fractal stimuli on both the RT and correctness in the oddball task, suggesting a strong effect of fractals on perceptual processing. Regarding the effects of attention to the stimuli in the oddball task, we did not find strong effects. Both fractal as well as non-fractal oddball stimuli show effects on both the correctness as well as the RT, indicating a general effect of attention (see figure 2). However, the effects on the RT and correctness of the stimuli go into different directions for fractals and non-fractals compared to the standard stimuli. Regarding these results, the 'fractal-rule learning' proposed by Martins et al. (2016) could also play a role here and explain the effect on higher correctness and lower RTs of the fractal stimuli as standards compared to fractals as an oddball. This effect entails, that after being presented with a few images, participants were able to infer an abstract principle or 'rule' behind the structures, which would then influence the results for the following images. Nevertheless, we observed a small effect that differed from the results of the standard, which might indicate some attentional effect of fractal stimuli. This, however, would need to be investigated in a future study, as the current results do not provide compelling evidence for that hypothesis. Specifically, creating a study in which one could exclude an effect of 'fractal-rule learning' could be very interesting.

In conclusion, fractal stimuli seem to have a strong effect on both reaction time as well as correctness, providing evidence for the special role of fractals in our processing pipeline. Moreover, processing of recursive structures is not limited to the linguistic domain. Fractal stimuli also seem to have a special role in attention, however, future studies are needed to investigate this further.

References:

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, *11*(3), 329–364.

Martins, M. D. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2055–2064.

Martins, M. D., Martins, I. P., & Fitch, W. T. (2016). A novel approach to investigate recursion and iteration in visual hierarchical processing. *Behavior Research Methods*, *48*(4), 1421–1442. https://doi.org/10.3758/s13428-015-0657-1.

Martins, M. D., Fischmeister, F. P., Puig-Waldmüller, E., Oh, J., Geißler, A., Robinson, S., Fitch, W. T., & Beisteiner, R. (2014). Fractal image perception provides novel insights into hierarchical cognition. *NeuroImage*, *96*, 300–308.

Supplementary Material:

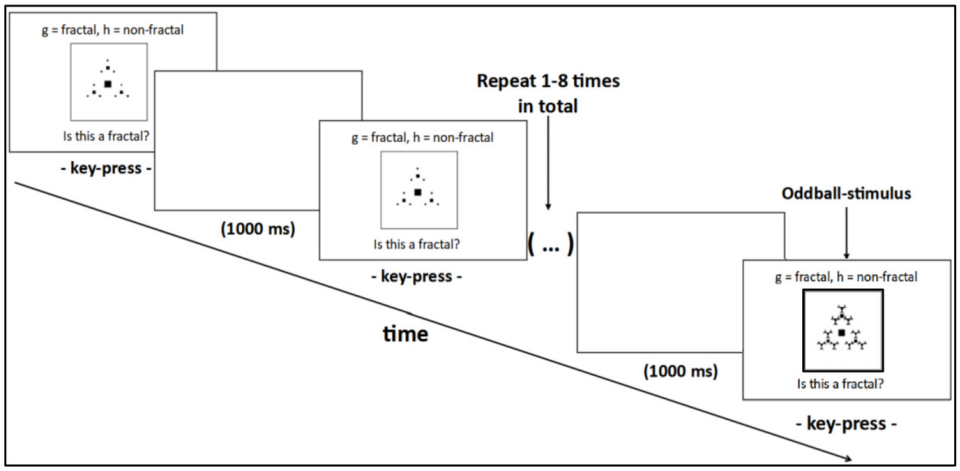


Figure 1: *Trial structure of the Oddball task (Example: Standard = fractal; Oddball = non-fractal). Participants were presented with one image in the centre of the screen and had to press either the G or H key to decide if the image was a fractal or a non-fractal. Mapping of responses to keys was counterbalanced.*

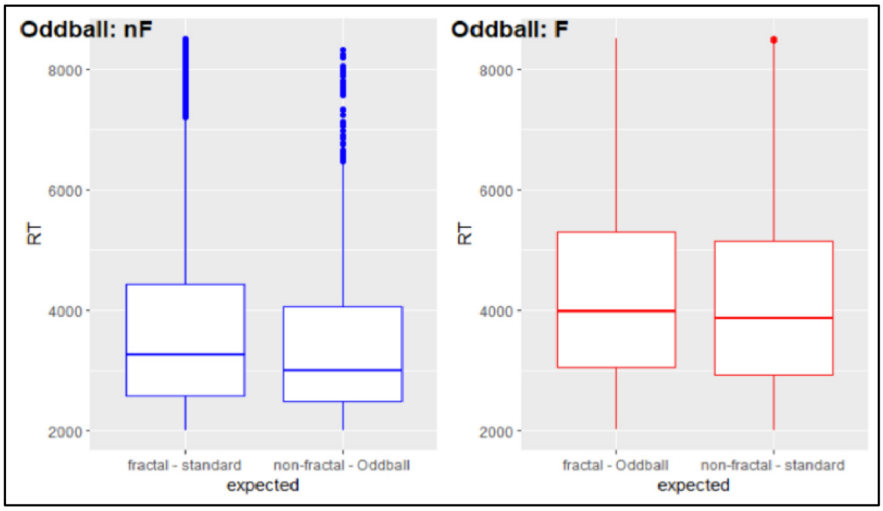


Figure 2: *RTs (in ms) the oddball task. The condition is indicated by the Oddball stimuli. ('Oddball: nF' = non-fractal (nF) Oddball and fractal (F) Standard; 'Oddball: F' = fractal Oddball and non-fractal Standard).*

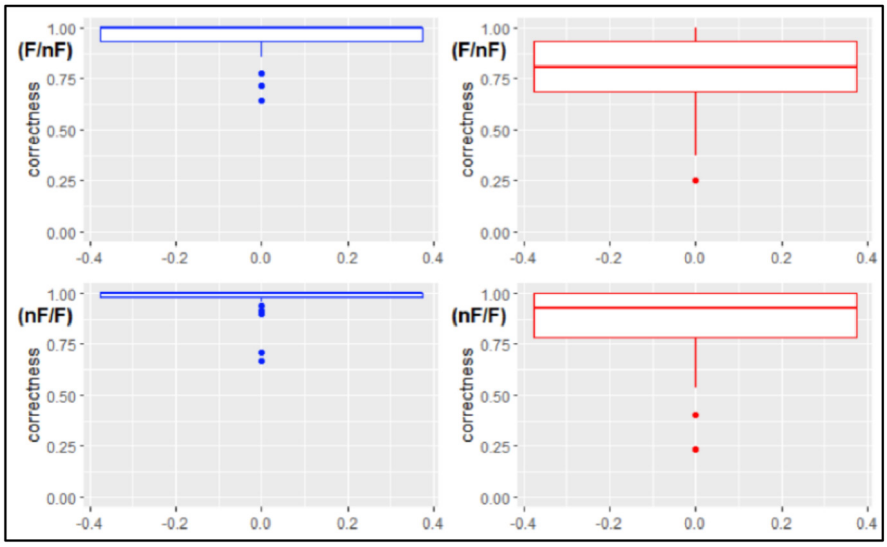


Figure 3: *This figure shoes the correctness of the different stimuli (blue = fractal (F); red = non-fractal (nF)) in the oddball task. The condition of the oddball task is indicated by the respective oddball and standard stimuli ((F/nF): Oddball = fractal; Standard = non-fractal | (nF/F): Oddball = non-fractal, Standard = fractal).*

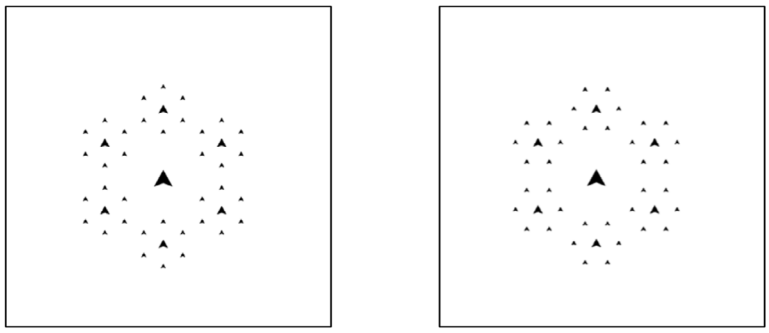


Figure 4: *Examples of fractal (left) and non-fractal (right) stimuli used in the oddball task. These specific examples are based on arrow-icons ordered in a hexagonal shape in different layers.*

Shepherd, Joshua (ICREA/Universität Autonoma de Barcelona) and Timothy Bayne (Monash University)

The capacity-based approach to functions of consciousness

Introduction

A foundational issue for the science and philosophy of consciousness concerns the function or functions of consciousness – what consciousness does for any particular aspect of psychological or neural processing. The issue remains contentious, with little agreement regarding the function of consciousness, whether there are multiple functions, how many functions there may be, and how different functions might relate to each other. One plausible reason for the lack of agreement is a basic need for more and better data. But in our view a better conceptual understanding of the issues surrounding the function of consciousness is required as well. Here we present the capacity-based approach to the functions of consciousness, and illustrate some of its virtues. In particular, we demonstrate how this approach can help us avoid stagnant ways of thinking about functions, can help us integrate results across disciplines, can shift explanatory priorities in a way that raises fruitful questions for consciousness science, and can help us work towards a more systematic, internally coherent understanding of the functions of consciousness.

Capacities and functions

The capacity-based approach flows out of a general explanatory approach that sees the mind as a structured collection of capacities. What capacities there are is a substantive theoretical question, but intuitive examples include the capacity to see, the capacity to reason, the capacity to attend, the capacity to remember, and so on.

Properly identifying the mind's key capacities, and charting how they relate to each other, is grounded in the further task of explaining how capacities operate. This is where functions enter into the approach. Functions are causal roles played by parts of a system, that enter into explanations of how capacities work, when they work.

Capacities, then, are explained by facts that identify functions (causal roles), and that enable us to lay out how it is that the capacity operates – how it is that some system performs the operation of seeing, or edge detection, or color discrimination, or reasoning, or inhibition, or task set construction, or attending, or remembering, or face perception, or emotion attribution, or whatever.

This approach to explaining capacities can be connected to a mechanistic approach to psychological and neuroscientific explanation, championed by Craver (2001, 2007), Krickel (2018), and Piccinini (2020), among others. On a mechanistic approach, a chief goal psychology and neuroscience share is the development of accurate descriptions of mechanisms – descriptions of the active, spatial, and temporal structure of systems of parts that illuminate how the joint operations of the parts explain causal outputs of the whole.

A capacity-based approach is a general approach to explaining the mind. Of course there will be significant uncertainty attached to current theorizing about many capacities. There is uncertainty attached

to any proposed mental ontology. But felicitously carving up the space of capacities is an ongoing, shared goal of the mind sciences, and is connected to the identification of the mechanisms and functional roles that explain the operations of capacities.

Consciousness

With a capacity-based approach to explaining the mind as background, the capacity-based approach to functions of consciousness poses the following as essential questions:

For any given capacity, does any aspect of consciousness play a causal role(s) in the operation of that capacity? If yes, what causal role(s)?

Notice, then, that the capacity-based approach relativizes claims about consciousness's function to capacities, as well as to aspects of consciousness. The relativization to capacities is important, for it is not a priori that the causal role played by consciousness for one capacity (e.g., color discrimination) will be at all similar to the causal role played by consciousness for another capacity (e.g., task switching, or error detection, or metacognition). And the relativization to aspects of consciousness is important for a similar reason. An aspect might be a particular feature of phenomenology (a particular type of phenomenal property), or a particular feature of the structure of a conscious state; different aspects of consciousness may play different functional roles for capacities.

Advantages

The capacity-based approach to functions of consciousness has several virtues. In this talk we highlight three.

First, in promoting a focus on causal roles within a mechanistic explanatory framework, this approach avoids mistakes endemic to the common strategy that asks, of any capacity, whether consciousness is necessary for the exercise of that capacity. Rather than a focus on necessity, we may do better to deploy a difference-making framework for understanding functions (Klein et al. 2020).

Second, the capacity-based approach need not commit to any particular theory of consciousness. Of course, if systematicity in the functional roles of aspects of consciousness begins to emerge, then this will lend itself to more sophisticated theorizing about the nature of consciousness. But the capacity-based approach can proceed independently from, or in conjunction with, speculation about the nature of consciousness. This theory-neutrality offers the prospect of progress on the functions of consciousness that side-steps thorny debates about consciousness's nature.

Third – and crucially, given the unsettled state of our knowledge of consciousness's functions – the capacity-based approach can re-orient explanatory priorities in a way that allows us to integrate findings across a disparate collection of research programs. We will illustrate this point briefly by discussing how best to interpret a recent study on the role of consciousness in certain visual detection effects (Stein and Peelen 2021).

Simpson, Valentine (University of Valencia)

Does Communication Involve Content? A Contention

A crucial point of contention concerning the phenomenon of (linguistic) communication is whether we should explain successful instances of it as a matter of subjects entertaining the same, or merely similar mental content in the exchange. Critics of same-content approaches claim it is too naïve a criterion, and give examples of successful communicative phenomena where sameness of content is undermined, and hence not necessary, such as in the communication of indexical thoughts, where their intimate connection to context make them not easily replicable (Recanati 2015). Some have gone even further, claiming that identity of content isn't even possible (Churchland 1998). In opposition, defenders of content sameness such as Fodor and Lepore have claimed that any robust notion of conceptual similarity will always presuppose a prior, more fundamental notion of conceptual identity (Fodor & Lepore 1992). Importantly, there is a widespread underlying commitment among these views that the fundamental notion by which we explain communication is that of content.

As a result of such heated debate, however, some have believed it worthwhile to explore an avenue devoid of all content, by which they attempt to explain communication with no reference to contents of any sort, or with requirements over and above content relations which purport to take their explanatory place. I call these "content neglecting" views. This is quite a revisionary view, as content has traditionally been assumed to play this key role in explaining communication, as well as other roles in the overall cognitive picture of the subject. In an ambitious piece, Pagin (2008) has discussed some views pertaining to this "family" head on, arguing that they are misdirected in their efforts: they more appropriately account for the general success of communication and communicative devices. They are explanations at "method level", Pagin claims (Ibid, p. 21), and not the "event level", which is the relevant level at which explanations of specific communicative episodes should aim.

I wish to engage with and, ultimately, refute one new contender for these content-neglecting views. Alexander Sandgren has recently developed (Sandgren 2019, 2021, 2022) a theory of intentional identity -the relation of a given number of beliefs having a common focus- that he calls the triangulation theory, which later feeds into his autonomist position -a radical view on mental and linguistic content- as a purported new account of communication. He proposes giving up explanations in terms of mental or linguistic content and adopting a new notion he calls "aboutness". Sandgren shifts focus from the object of a belief to the relations held with other representations. Aboutness is thus understood in meta-representational terms, explained via the "agents' beliefs about what it would take for attitudes to be about the putative target" (Sandgren, 2019, p. 3685). A subject's second-order beliefs are appealed to in characterising the aboutness of some first-order belief. These second-order beliefs lay out "triangulation conditions" which capture the fact that any representation targets the same as another when it has been formed in a sufficiently similar way, or represents the same target in a sufficiently similar fashion. Another subject's belief is about the same when it satisfies at least one of its triangulation conditions.

This triangulation theory is then conjoined with a particular (and, importantly, highly revisionary) view on mental and linguistic content which he dubs autonomism. This view claims that we cannot recover the content of a belief from the content of a sentence which expresses it, nor vice-versa. Mental content and linguistic content are autonomous from each other, answering to different constraints. Mental content, he argues, is absolutely sensitive to the agent's individual psychological states, while linguistic content isn't, as it seems to answer to some more public or external factors. His crucial thesis, then, is that communication should be explained in terms of aboutness and not content. Where others explained cases of communication by positing sameness/similarity of content, Sandgren looks to the co-aboutness of the beliefs. Communication could involve different mental contents for the parties involved as long their beliefs triangulate on the same objects and properties.

I challenge Sandgren's content-neglecting framework on the grounds that he cannot account for the relevant phenomena only in terms of his own theory/notion, with no reference to content. Taking aboutness (understood in this way) as a constitutive feature of beliefs leads to some problems. We need some ground-level of content, for, if not, what must the triangulation theory say about the subject matter of these higher-order beliefs when they serve as the basis for our assessment of another belief's subject matter? This lands Sandgren into a dilemma: If we are to find some further (even) higher-order belief to capture its aboutness, we are launching into an infinite regress. If we appeal to some belief we have already appealed to at some point in the chain, we have a disturbing circularity. This shows we need some element for the characterisation of our propositional attitudes that doesn't appeal to other, further attitudes; we need something serving as the basis or building blocks, something exactly like content.

In the spirit of Pagin's earlier criticism, the idea is that the fact that we may draw relations of identity between objects involved in communicated beliefs through the higher-order beliefs subjects (might) have about them is good evidence that communication, in general, succeeds. We want these meta-representational beliefs, when available, to guide us to the things the subjects think about and express. However, this only gets us an explanation at method level, a broad guide as to how connections in the subject's beliefs connect to communicative practices; there still needs to be some ground-level explanation of specific instances of communication (event level) in terms of the involved subjects' mental contents, or we are left with faulty explanations. This, I take it, is the main lesson to take away from both Pagin's and my discussion of this family of content-neglecting views: unless we appeal to content, we are left without a key ingredient in our explanations of specific episodes of communication.

References:

Churchland, P. M. (1998). Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered. *The Journal of Philosophy*, *95*(1), 5–32. https://doi.org/10.2307/2564566

Fodor, J. & Lepore, E. (1992). *Holism: A Shopper's Guide*. Blackwell.

Sandgren, A. (2019). A metarepresentational theory of intentional identity. *Synthese* 196 (9): 3677–3695.

Sandgren, A. (2021). Turning Aboutness About. *Pacific Philosophical Quarterly* (1):136–155. https://philarchive.org/archive/SANTAA-5v2

Sandgren, A. (2022). Private Investigators and Public Speakers. *Journal of the American Philosophical Association*, 1–19. doi:10.1017/apa.2021.50

Pagin, P. (2008). What is Communicative Success?. *Canadian Journal of Philosophy*. 38. 85–115. 10.1353/cjp.0.0011.

Recanati, F. (2015). Indexical Thought: The Communication Problem. In Manuel Garcia-Carpintero & Stehpan Torre (eds.), *About Oneself*. pp. 141–178.

Skoczeń, Izabela (Jagiellonian University)

Lies, Common Ground and the Law

There are two famous US cases concerning the legal definition of a lie (perjury). In both cases the defendant said something literally true, but only one was a conviction (DeZarn v US), while the other (Bronston v US) an acquittal. Bronston said something true that implicated something false, while DeZarn answered truthfully a question about an irrelevant fact and didn't correct the inquiring lawyer. The differing judicial decisions can be explained by the linguistic common ground theory (see for instance Stalnaker, 1978; Pagin, 2016). This is a theory, which claims that the common ground is a set of propositions that the interlocutors agree to treat as true and each new contribution to the conversation is a proposal to update the common ground, or to treat the additional proposition as true. In the legal cases only Bronston tried to update the common ground and thus it was the duty of the lawyer to inquire further if Bronston's implicature was true or false. However, in both cases the jury stated that the defendant lied and is guilty. In the experiment (N=262) I presented participants with a vignette describing John, a candidate for local governor, who attended two parties, one in the spring and one in autumn; at one of the parties he gave bribes to the guests so that they would vote for him in the elections. Next, participants were randomly assigned one of the three conditions below (labels omitted):

1 Question: The questioning lawyer asks John whether he gave bribes to the guests at the autumn party. John replies:

'I did not give any bribes at the autumn party.'

2 Implicature: The questioning lawyer asks John whether he gave bribes to the guests at any party. John replies:

'I did not give any money to anyone at any party.'

3 Control: The questioning lawyer asks John whether he gave bribes to the guests at any party. John replies:

'I did not give any bribe to anyone at any party.'

Next, participants were asked whether John intended to make part of the common ground that he gave no bribes at any party. Subsequently, participants were informed that: John gave bribes in the form of expensive delicacies at the spring party but not at the autumn party.

Finally participants had to rate how much they agreed with the statement that John's answer is a 'real lie, even though technically it is not a lie'.

I find that lay participants do not perceive the difference in common ground updating and agree that the speaker's utterance 'is a real lie, even though technically, it is not a lie' in all conditions. The answers on both dependent variables are correlated. I conclude that this might be due to the fact that the core of the folk concept of a lie is the intent to deceive, while the legal definition is different, as it revolves around the maxim that no one has the duty to self-incriminate oneself.

Skrzypulec, Błażej (Jagiellonian University)

Experiential Boundaries and the External Space

A common idea proposed in the context of analyzing perceptual experiences is that perceptual experiences possess certain structures (Alsmith, 2017; Macpherson, 2015; Phillips, 2013; Richardson, 2010; Soteriou, 2013). An important category of perceptual structures are limitation structures which determine the limits of sensory experiences. My talk concerns a particular type of spatial limitation structures which I call 'boundary structures'. These are structures which demarcate the limits of perceptual space in which entities can be perceived. For instance, in vision the boundaries of the visual field determine a space in which objects can be seen.

More precisely, I focus on one particular thesis, which I name the 'Externality Thesis', regarding the boundary structures. It is claimed, referring both to the visual and bodily experiences, that in virtue of awareness of perceptual boundary structures, we are aware that the visual or bodily space is a part of a larger space (Cavedon-Taylor, 2018; Dokic, 2003; Laasik, 2019; Mac Cumhaill, 2015; Martin, 1992, 1993; Richardson, 2010, 2013; Serrahima 2022; Soteriou, 2013; Wilson, 2022). My goal is to evaluate the Externality The in the context of vision and touch. More specifically, relying on mereotopological theories (Casati and Varzi, 1999; Smith, 1997), I show that the notion of spatial boundaries is ambiguous as it encompasses various distinct ways in which entities may be limited. I argue that only some of these ways are able to support the Externality Thesis. In particular, I claim that while bodily boundaries of which one is aware in tactile sensations are such that the awareness of them provides a support for the Externality Thesis, the analogous claim is not true about boundaries of the visual field.

In order to argue for this position, I distinguish, relying on mereotopological considerations, two types of boundary structures. The first type provides the 'external support' for the Externality Thesis. In the case of such structures, the awareness of the perceptual space and its boundary is not enough to support the Externality Thesis as what is additionally required is the awareness of an entity located outside the boundary of a perceptual space. The situation is different in the case of boundary structures which provide the 'internal support'. If a boundary structure provides the internal support for the Externality Thesis, the awareness that there is something beyond the perceptual space does not require an awareness of something that is positioned outside the perceptual boundaries. What is sufficient is an awareness of a perceptual boundary itself.

Relying on the above distinction, I argue that bodily boundaries present in typical tactile sensations are such that they provide an external support for the Externality Thesis. Due to the bipolar character of touch (Mattens, 2016; Ratcliffe, 2008; Richardson, 2013), in the case of many tactile sensations, awareness of a boundary structure is an awareness of a boundary between the bodily space and an object which possesses some spatial property. The boundary structure in which a boundary separates bodily space from an object having spatial properties conveys information that there is an object which is located somewhere, due to the fact that it has some spatial properties, and its location does not overlap with the bodily space. This means that beyond places composing the bodily space, there is also at least one other place. Such support for the Externality Thesis is an external support as the crucial piece of information is provided by the awareness of something positioned outside the bodily space.

On the other hand, I argue that the awareness of the boundaries of the visual field provides neither external nor internal support for the Externality Thesis. First, relying on the phenomenal fact that we are not visually aware of anything what lies beyond the boundaries of the visual field, I argue that there is no element of the visual boundary structure whose awareness could provide an external support for the Externality Thesis. In consequence, if visual boundary structure provides support for the Externality Thesis, then such support has to be internal.

Second, I argue that to provide internal support for the Externality Thesis, boundaries of the visual field have to be such that the awareness of them is also a perceptual awareness that these boundaries overlap or coincide with something distinct from visual space. Further, I show that such perceptual awareness can happen only in virtue of the mechanisms of peripheral, and not central, vision. Relying on that, I argue that it is unlikely that the peripheral vision can provide an awareness that visual boundaries overlap or coincide with something distinct from visual space. In particular, being aware of boundaries overlap or coincidence

requires making a precise recognition of relations in which fragments of the visual space stand. In particular, it requires recognizing that a part of the visual space stands in a certain parthood relation to something distinct from visual space or that there is something distinct from visual space which is co-located with this part. Such recognition is difficult in the case of peripheral vision which is not able to precisely discriminate relations between places and spatially instantiated features.

Furthermore, the difference between awareness of the space in the center of the visual field and the awareness of space at the edges of the visual field does not seem to be distinct in some positive way. It is rather merely a negative difference consisting in the fact that at the edges the awareness of the structure of visual space is less detailed. Because of that it is unlikely that in the awareness of the edges of the visual field we are aware of some additional relation regarding overlap or coincidence. In comparison to the awareness of the central space, the awareness of peripheral space is diminished and not enriched by some additional element indicating the presence of something distinct from the visual space.

Sources

Alsmith, A. J. T. (2017). Perspectival structure and agentive self-location. In F. de Vignemont, & A. J. T. Alsmith (Eds.), *The subject's matter: Self-consciousness and the body* (pp. 263–287). Cambridge, MA: The MIT Press.

Casati, R., & Varzi, A. (1999). *Parts and places. The structures of spatial representation*. Cambridge, MA: The MIT Press.

Cavedon-Taylor, D. (2018). Sensorimotor expectations and the visual field. *Synthese*, 198, 3991–4006.

Dokic, J. (2003). The sense of ownership: An analogy between sensation and action. In J. Roessler (Ed.), *Agency and self-awareness: Issues in philosophy and psychology* (pp. 321–344). Oxford: Clarendon Press.

Laasik, K. (2019). Visual field and empty space. *European Journal of Philosophy*, 27(2), 403–411.

Mac Cumhaill, C. (2015). Perceiving immaterial paths. *Philosophy and Phenomenological Research*, 90(3), 687–715.

Macpherson, (2015). The structure of experience, the nature of the visual, and type 2 blindsight. *Consciousness and Cognition*, 32, 104–128.

Martin, M. (1992). Sight and touch. In T. Crane (Ed.), *The contents of experience* (pp. 196–215). Cambridge: Cambridge University Press.

Martin, M. (1993). Sense modalities and spatial properties. In N. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 206–218). Oxford: Oxford University Press.

Mattens, F. (2016). The sense of touch: From tactility to tactual probing. *Australasian Journal of Philosophy*, 95(4), 688–701.

Phillips, I. (2013). Hearing and hallucinating silence. In F. Macpherson, & D. Platchias (Eds.), Hallucination. *Philosophy and psychology* (pp. 333–359). Cambridge, MA: The MIT Press.

Ratcliffe, M. (2008). Touch and situatedness. *International Journal of Philosophical Studies*, 16(3), 299–322.

Richardson, L. (2010). Seeing empty space. *European Journal of Philosophy*, 18(2), 227–243.

Richardson, L. (2013). Bodily sensation and tactile perception. *Philosophy and Phenomenological Research*, 86(1), 134–154.

Serrahima, C. (2022). Sensory fields: the visual and the bodily. *Philosophical Studies*, 180, 679–700.

Smith, B. (1997). Boundaries: An essay in mereotopology. In L. Hahn (Ed.), *The philosophy of Roderick Chisholm* (pp. 534–561). LaSalle: Open Court.

Soteriou, M. (2011). The perception of absence, space, and time. In J. Roessler, H. Lerman, & N. Eilan (Eds.), *Perception, causation, and objectivity* (pp. 181–206). Oxford: Oxford University Press.

Soteriou, M. (2013) *The mind's construction*. Oxford: Oxford University Press.

Wilson, K. A. (2022). The auditory field: The spatial character of auditory experience. *Ergo*, 9, https://doi.org/10.3998/ergo.2909.

Stamatiou, Filippos (University of Copenhagen)

Moral luck and "many-shots" actions

Moral luck suggests that our actions are prone to luck. The possibility of moral luck indicates that we cannot be morally responsible for our actions and their outcomes. Thus, theories of moral responsibility are forced to adopt some version of internalism, or resort to scepticism. I argue that prominent cases of moral luck are founded on the assumption that actions are events. I propose an alternative ontology of action as an ongoing goal-directed process with a "many-shots" structure that extends over time. Described this way, cases of moral luck are not representative of ordinary action, and therefore, not generalisable. Thus,

the threat that luck poses for theories of moral responsibility is diminished significantly. This proposal of actions as a many-shots process is consistent with predictive coding, a framework for cognitive architecture which centres around error minimisation. Under this framework, cases of resultant more luck are no longer failures of action, but rather anticipated errors to be minimised within the ordinary process of action.

If control is a necessary condition for moral responsibility, one cannot be responsible for what is beyond one's control. Luck refers to factors that beyond one's control. Moral luck in particular refers to factors beyond one's control that appear to make a moral difference. This suggests a conflict. How can luck make a difference in one's degree of moral responsibility when it is beyond one's control? Either control is not a necessary condition for moral responsibility, or luck only appears to make a moral difference. I provide support for the second option. In this paper, I argue against the claim that because of apparent cases of moral luck, people are not morally responsible for actions and their consequences.

The paper is a contribution to the growing literature on the possibility of a naturalistic account of free action in light of serious challenges, one of which is luck. I am concerned in particular with one kind of luck, namely resultant moral luck. I do not provide an argument against resultant luck in general. Rather, I object to a certain use of apparent cases of resultant luck, which have been used to claim that people are not morally responsible for their actions and their consequences. Thus, resultant moral luck seems to force a choice upon us: One must either adopt a sceptical position or restrict the scope of moral responsibility to the internal mental life of the agent. It is exactly to this use of the argument that I object.

I begin by introducing moral luck in general, before I move on to a discussion of prominent responses. As most available accounts demonstrate, it is hard to explain away the influence of resultant luck on an agent's blameworthiness. In fact, after considering typical cases of resultant moral luck, one is pushed to one of two unattractive alternatives: internalism or scepticism. I discuss why both alternatives are unattractive, before considering prominent of resultant moral luck in more details. I argue that they are built on an assumption that is rarely made explicit, namely that actions are events. I discuss the implications of that assumption and argue that there might be a better way to capture ordinary action.

Then, I present such an account. I provide an alternative ontology of action, one that departs from the "one-shot" event structure and moves towards describing action as an ongoing goal-directed process with a many-shots structure. The process ontology suggests that actions are gradually unfolding over time, subject to change, and best captured in progressive sentences. Described under the alternative ontology, instances of resultant moral luck appear as special cases, no longer representative of ordinary action. While this does not exclude the possibility that luck may still influence an agent's blameworthiness, it significantly diminishes the force of the challenge by resultant moral luck. Thus, the credibility resultant moral luck provides to the view that people are not morally responsible for their actions is also diminished.

Finally, I examine whether there is a good candidate for a cognitive framework which can provide support for the process ontology of action. I turn to predictive coding as a plausible, naturalistic proposal for human cognition. I argue that predictive coding is consistent with the many-shots structure of action. Within a predictive coding framework, cognition is primarily engaged in error minimisation. Errors like the ones captured in cases of resultant luck are not necessarily failures, but rather integral elements of the many-shots process of action. Thus, predictive coding provides support to the claim that cases of resultant moral luck do not show that people are not morally responsible for their actions and their consequences. I finish with a discussion of potential objections, before considering a proposal for how to understand apparent cases of moral luck.

Stazicker, James (King's College London)

Perceptual confidence, presence to the mind, and disciplinary autonomy

The view that perceiving involves various degrees of confidence in hypotheses about one's environment plays a significant role in several recent scientific, philosophical and interdisciplinary theories of perception and perceptual experience (Friston et al. 2006, Clark 2013, Howhy 2013, Seth 2014, Morrison 2016, among others). This talk starts by defending a negative claim: such perceptual confidence is not fundamental

to a good constitutive account of the subjective character of perceptual experience; some degree of confidence in a hypothesis may sometimes figure in the subjective character of perceptual experience, but when this occurs it depends on distinct, more basic ways of experiencing one's environment. The talk then exploits the example of perceptual confidence to develop a positive proposal about the autonomy of philosophical and scientific theories of perception from one another. In short, notions of representation in the philosophy and science of perception are often best understood in terms of different explanatory projects, such that philosophical and scientific commitments about perceptual representation are neither equivalent to, nor directly supported by, one another, even when they are applied to the same episodes of conscious perceptual experience.

To articulate a neutral constraint on good constitutive accounts of the subjective character of perceptual experience, we begin outside contemporary representationalist frameworks, with the traditional idea that in perceptual experience things are 'present to the mind'. Hume operationalised this in terms of demonstrative linguistic expressions: 'the existences, which we consider, when we say, this house and that tree' (Hume 1777: 12.1; Snowdon 1992). Although Hume argued that the objects satisfying this criterion for presence to the mind are impressions, rather than the mind-independent tables, trees etc we unreflectively assume, his operationalisation of this central aspect of the subjective character of perceptual experience is independent of that argument. To extend the operationalisation to perceptual experience of features, rather than objects, we can exploit parallel linguistic expressions such as 'that shape', 'that colour', etc. In common with the demonstrative expressions referring to perceived objects, these latter expressions have meanings which are grasped only through perceptual experience of what is demonstrated (Peacocke 1989). A feature of the environment is present to the mind, in the way characteristic of perceptual experience, only if the experience puts one in a position to entertain thoughts about this feature which are canonically expressible in this demonstrative way.

Morrison (2016) argues that perceptual experience constitutively involves various degrees of confidence in hypotheses about one's environment, on the grounds that this makes room for the immediate justification, by perceptual experience, of varying doxastic credences in propositions about one's environment. For example, take a typical subject in a typical psychophysical experiment, whose comparative judgements about the shape of a stimulus reveal a bell-shaped curve of doxastic credences over possible exact shapes of the stimulus: the subject's credences in various oval shapes are relatively high, say, while her credences in rounder-than-oval shapes are progressively lower as they approach circularity, and her credences in thinner-than-oval shapes are progressively lower as they tend further from circularity. According to Morrison, in order to respect the immediacy of perceptual justification, a constitutive account of perceptual experience should postulate a correspondingly bell-shaped distribution of perceptual confidences over possible shapes of the stimulus. In particular, such perceptual confidences are better placed to respect the immediacy of perceptual justification than is the outright perceptual attribution of a single determinable or unspecific shape to the stimulus.

On the contrary, I argue, perceptual confidences over possible shapes cannot accommodate the immediate justification, by perceptual experience, of judgements which operationalise the central aspect of the subjective character of perceptual experience described above. This is because, unlike credences over possible shapes, perception-based judgements canonically expressible using the phrase 'that shape' license outright commitment to the shape's presence in the environment. To respect this, we should characterise perceptual experience of the shape in terms of the outright perceptual attribution of a single determinable or unspecific shape to the stimulus.

Doxastic credences over possible exact shapes of the stimulus may show up in how things look to a subject, broadly construed, but the credences are not reflected by corresponding structure in the representational contents of perceptual experience. Yet productive scientific models of perception turn on the assumption that perceiving involves various degrees of confidence in hypotheses about one's environment. How can this be?

In general, scientific and philosophical projects often make prima facie contradictory claims about perception. For example Gregory (1980) maintained, and McDowell (1994) denied, that perceptions are hypotheses. Following McDowell, a standard way to reconcile such claims proposes that they target distinct psychological states: what Gregory described as hypotheses are states of an information-processing subsystem within an organism; what McDowell denied are hypotheses are conscious perceptual experiences of a whole person. However, this route to reconciliation is untenable in the face of contemporary scientific work which targets the experiences of a whole organism (Friston et al. 2006; Seth

2014); this work's explanatory power depends on treating conscious perceptual experiences of the whole organism as hypotheses.

A better route to reconciliation draws the personal / subpersonal distinction not as a distinction between different psychological states, but as a distinction between different types of psychological explanation (see Drayson 2012). For example, if perceptual experiences are hypotheses, this is so in virtue of their explanatory role in a system of computations postulated by scientific psychology; by contrast, perceptual experiences do not have the status of hypotheses in the context of the rational explanations of action or belief which characterise a subject's point of view. Thus it may be literally true both that the same states or episodes of perceptual experience are hypotheses and that they are not hypotheses, given two distinct notions of 'hypothesis' defined by two distinct explanatory projects. As a result, philosophical and scientific theories which treat perceptual experience as a form of representation should, by and large, draw only indirectly on one another when specifying the types of representation involved, at least where philosophical theories concern the subjective character of experience.

Anna Strasser (Ludwig-Maximilians-Universität München)

What shall we do with the increasing indistinguishability between machines and humans?

Artificial agents are becoming increasingly prevalent in human social life. In many situations, the perceivable difference between artificial systems and human actors is getting blurred. This concerns not only mere machine-generated outputs but also the way humans interact with machines. Regarding machine-generated content, we are exposed, for example, in social media, e-customer service, and advertising, to content that can easily be mistaken for human-generated content. In particular, the rapid development of large language models (LLMs) has highlighted that neither humans nor sophisticated detection software can distinguish with certainty between human-generated and machine-generated text. Indeed, with the rise of social robotics, humans are cultivating the ability to form emotional bonds with inanimate artificial agents, even treating artificial agents as if they were social agents. For instance, it has been shown that humans can sometimes feel committed to robot joint action partners. All this might question approaches that describe all human-machine interactions as mere tool use and not as genuine social interactions. And at the same time, this raises far-reaching questions, namely whether artificial systems should also be assigned moral agency and the corresponding rights and obligations.

The talk investigates challenges such as no longer undoubtedly verifiable human authorship, a new form of plagiarism, or the massive spread of misinformation as well as opportunities that the use of LLMs in everyday life might have.

Tooming, Uku (University of Tartu) and Kengo Miyazono (Hokkaido University)

Epistemic Justification through Memory Consolidation

Memory is plausibly seen as a preservative source of epistemic justification in that its role is to transmit justification from other sources like perception or reasoning. Recently, this view has been questioned from various sides, and it has been proposed that memory can generate justification on its own (Michaelian 2011; Lackey 2005, Fernandez 2016, Boyle 2019).

In this paper, we will propose another argument for the view that memory is epistemically generative. Our argument relies on empirical evidence on memory consolidation which, as far as we are aware, has not been considered in the epistemology of memory literature.

Our definition of epistemic generativity of memory goes as follows.

Memory is epistemically generative iff it is not the case that, for all subjects, S, propositions, P, if S is justified to believe that P by J1 at a time T2, then it is because, at an earlier time, T1, S was already prima facie propositionally justified to believe that P by another source of justification, J2.

As we have argued in our earlier work, this definition of epistemic generativity is appropriately demanding and most proposed views of generationism about memory do not satisfy it [redacted]. The definition is appropriately demanding because it is not sufficient for a case of memory to count as epistemically generative if memory justifies a belief formed on its basis. The justification in question also has to be such that it is not explained in terms of prima facie propositional justification by some other source at an earlier time. If that requirement is not met, it is always possible for an epistemic preservationist about memory to argue that, in general, either a case of putative memory justification is such that the justification that S has can be explained by the fact that S was justified to believe that P already at an earlier time by some other source or S is not actually justified to believe that P in that case.

Because of these considerations, we can say that generationism about memory requires the existence of generative cases of memory that satisy the following conditions:

JUSTIFICATION: S is prima facie propositionally justified to believe that P by memory at T2.

NOPRIOR: There is no earlier time, T1, at which S was already prima facie propositionally justified to believe that P by another source of justification.

What we argue in this paper is that there are cases of memory that satisfy these two conditions and that can be found when we consider the empirical studies on memory consolidation. As a result, memory can also satisfy the conditions for epistemic generativity in the demanding sense.

"Memory consolidation" refers to post-encoding mnemonic processes during which memories are stabilized and integrated with the agent's preexisting memories (Nadel & Moscovitch 1997; Klinzing, Niethard & Born 2019). Here, we focus on sleep-dependent consolidation processes in particular (Stickgold 2005). What is crucial for our purposes is that memory consolidation during sleep can change the subject's epistemic position with respect to some proposition, P, such that the subject is prima facie propositional justified to believe that P on the basis of memory (JUSTIFICATION), while she was not prima facie propositionally justified to believe that P before memory consolidation by some other source (NOPRIOR).

Consolidation can make a difference to the subject's epistemic position because memory systems transform memory content during the consolidation phase. The most salient process during memory consolidation is schematization and loss of specific contextual detail of represented events (Winocur & Moscovitch 2011). However, this process also enables the extraction of abstract regularities from past experiences that can then be generalized for future situations (Cowan et al. 2021). By extracting those regularities, memory consolidation process can also put the subject in a position to form beliefs about those regularities that she was not able to form before going through that process.

For example, in a study by Wagner et al. (2004), the participants had to convert an 8-digit string into a new string by a stepwise application of two simple rules. A hidden rule was applicable to those transformations but the participants were not informed about it. Sleep increased more than twofold the probability of gaining insight about the rule in question (for comparable data on how sleep improves the capacity to learn implicit rules, see Battering et al. 2014). It is plausible that for at least some of the participants, sleep-dependent memory consolidation was crucial in making it the case that they acquired justification to form beliefs about the rule in question. Furthermore, that they were not able to comprehend the rule before sleep suggests that they were not in the position to form beliefs about the rule before the consolidation processes. If they were not in that position, it is also plausible that they were not prima facie propositionally justified to form beliefs about the rule before memory consolidation. Consolidation thus changes the subject's epistemic position with respect to propositions about those rules.

Imagine Jane who engages in the kind of task that was presented in Wagner et al. At T1, before sleep-dependent memory consolidation, she is not able to identify the rule in question and is therefore not in the position to justifiably believe that she performed a task to which such a rule applied, i.e., she is not prima facie propositionally justified to believe that she performed such a task. However, over sleep, the memory consolidation processes extracted that rule from the representation of the task and, as a result, at T2, she is in a position to justifiably believe that she performed a task to which the rule applied. Jane's case satisfies JUSTIFICATION and it satisfies NOPRIOR: it is not the case that the justification by memory that she has at T2 is due to justification by some other source at T1.

Toribio, Josefa (ICREA-UB)

Seeing wrongness

It is a fact that some of our actions strike us as being morally charged, that we can be aware of the moral significance of some of the things that we say or do. It is also a plain fact that we often fail to do so. When we notice the wrongness of an action, whether the action is ours or someone else's, what is the nature of this recognition, and what does it tell us about how we come to know about such a moral property? A proposal that has regained some popularity in addressing this issue is moral perceptualism (e.g., Audi 2013, Cullison 2010, McBrayer 2010, McGrath 2011, Werner 2016). If moral perceptualism is true, we can perceive instantiations of at least some moral properties—e.g., descriptively 'thin' evaluative properties, such as being right/wrong or being just/unjust—in a way that is comparable to the way in which we perceive other complex properties. According to this view, which following Vance and Werner (2022), I shall label Contentful Moral Perceptualism (CMP), agents can represent (some) moral properties as part of the content of their perceptual experience.

CMP has been challenged on several fronts. It has been argued that perception traffics in causal connections, and it is hence ill suited to represent moral properties, since moral properties are causally inert (see e.g., Griffin 1996, Huemer 2005). Another point of criticism has been the absence of a characteristic perceptual appearance or "look" of moral properties (see e.g., Reiland 2021). Relatedly, critics have also raised the issue of the explanatory redundancy of perception given that perceptual experience would not be a morally charged experience without prior moral beliefs or knowledge (see e.g., Cowan 2015, Faraci 2015, Reiland 2021). Despite some recent rejoinders to these objections (see e.g., Cullison 2010, McBrayer 2010, Werner 2018), the claim that we can represent moral properties in perception remains a very contentious one.

More recently, the idea that we can perceive moral properties has been characterized as consisting in a heightened tendency to attend to morally salient features, without this entailing a commitment to the idea that moral properties can be part of the content of perception. Vance and Werner (2022) call this view Attentional Moral Perceptualism (AMP) (see also Clifton 2013, DesAustels 2012). AMP's central claim is that the pre-theoretical intuition that some of our actions strike us as right or wrong can still be explained by the sensitivity of certain perceptual, albeit attentional, mechanisms to "moral difference-makers", which "are the features within one's perceptual environment that make a moral difference in that situation" (Vance and Werner, 2022, p. 9). By shifting the explanation from content to attention, AMP appears to be a more empirically and philosophically plausible version of moral perceptualism—a version that is free of standard objections against CMP.

This paper discusses the plausibility of Attentional Moral Perceptualism (AMP) vis-à-vis our failing to see the wrongness of actions that result from our implicit biases. The strategy is thus to discuss cases in which the property of an action being wrong is overlooked to shed light on how we come to know about such a property. First, I challenge that the empirical evidence offered to support AMP does indeed support any *perceptual* attentional sensitivity to morality. Most of the results, I argue, are better explained as involving cognitive phenomena, such as semantic priming and negativity bias. Attention does enable (and prevents) perception, but it involves both perceptual and cognitive functions. Second, I argue for the following dilemma: on the one hand, if attention leads to changes in the content of perceptual experience, then AMP suffers from the same problems that afflict its content-based predecessors. On the other hand, if attention does not lead to changes in representational content, then AMP will have to resort to some specific attentional phenomenology that runs independently of the content of perceptual experience. But the very nature of such an attentional phenomenology is very difficult to pin down, and, at least on an intentionalist interpretation, the notion leads to very counterintuitive consequences regarding the transparency of perception, as it suggests that when attending, our experience is about the properties of such an experience instead of about the properties represented by the experience.

More positively, I argue that attention does play an important role in our recognition of moral properties, but that AMP only kicks the explanatory bucket further, by opening up the question of what drives our attention when facing morally charged situations. By putting together elements from the biased competition view of attention and a recent account of prejudice defended by Jessie Munton (2021), I finally

suggest an account of our (in)sensitivity to wrongness that incorporates the important role that attention plays in AMP without a commitment to there being either perceptual representations of moral properties or perceptual, attentional mechanisms that are specifically responsive to morality.

Vanello, Daniel (Philosophy Department, University of Warwick)

Moral identity, moral integration, and autobiographical narrative

The aim of this paper is to assess the claim made by a number of moral identity theorists that autobiographical narratives establish the presence of moral integration. I argue that this claim is unwarranted. I argue that the presence of moral integration is established by methods of assessment independent of autobiographical narrative.

The concept of 'moral identity' has become prevalent in research in moral development (Hardy, Krettenauer and Hunt, 2020). The concept of moral identity was originally introduced by Augusto Blasi to explain the connection between moral judgement and moral action (Blasi, 1984). One of the key claims of moral identity theorists is that one is motivated to act according to their moral judgement when morality is important, or central, to one how defines oneself. This is captured by a commonly quoted definition of moral identity as 'the degree to which being a moral person is important to an individual's identity' (Hardy and Carlo, 2011, 212).

Moral identity theorists conceptualise the idea of morality becoming central to how one defines oneself in terms of "moral integration". Moral integration refers to the process by which morality becomes part of one's self-definition.

A number of moral identity theorists have appealed to the concept of autobiographical narrative to study moral integration and its role in the explanation of moral action (e.g. Krettenauer and Mosleh, 2013). Autobiographical narratives are stories told about oneself that render intelligible different aspects of one's self-definition. Autobiographical narratives appeal both to one's past experiences and semantic knowledge about oneself (Fivush, 2020).

In this paper, I distinguish between three claims that moral identity theorists often conflate:

Claim 1: autobiographical narratives identify the presence or absence of moral integration in an individual.

Claim 2: autobiographical narratives predict moral action.

Claim 3: autobiographical narratives explain the process of moral integration.

In this paper, I focus on claim 1. I argue that Claim 1 is unwarranted.

My argument focuses on one of the most important methods to establish moral integration in an individual used by moral identity theorists. That is, Frimer and Walker's reconciliation model of moral integration (Walker and Frimer, 2015). Their model argues that moral integration is the result of reconciling a conflict between two motivational systems they call agency and communion. Agency refers to the motivational system constituted by self-interested goals. Communion refers to the motivational system constituted by altruistic goals. According to the reconciliation model, moral exemplars reconcile the agency and communion motivational systems. Crucially, Frimer and Walker introduce a new method to study reconciliation: the values embedded in narrative (VEINs) method. VEIN's involves three parts. First, participants are asked a set of questions that are meant to elicit autobiographical narratives, that is, stories about oneself that connect past events in one's life, including past experiences and semantic knowledge about oneself, into a coherent personal biography. The second part appeals to a number of coding systems used in other settings that are meant to identify the moral values, if any, that the participants appeal to in their autobiographical narratives. These coding systems are then used to argue whether or not a given participant's autobiographical narrative is constructed around a moral value. If so, then the participant is scored highly on the moral centrality scale. In turn, according to this method, the participant exhibits moral integration. In essence, VEINs assumes that moral integration is present if the autobiographical narrative is deemed as appealing to both the agentic and communion motivational systems without showing any conflict between the two. According to Frimer and Walker, this is an example in which the participant

integrates agency and communion because it demonstrates that the participant is motivated to act according to certain moral values that they see as central to how they define themselves. In the third part of the study, participants were assessed on their moral functioning. This part asked the participants about the frequency with which they engaged in moral behaviour, specifically prosocial or altruistic behaviour, ecological behaviour and whether they engaged in behaviour that prized material possessions over altruism. The responses were then scored from high to low. According to Frimer and Walker, VEIN's showed strong correlations between scores from part two and part three of the study thus providing evidence that autobiographical narratives are correlated to moral integration.

I argue that the reconciliation model does not show that autobiographical narratives establish the presence of moral integration. Frimer and Walker (2009) assess participants on their moral functioning by asking the participants about the frequency with which they engaged in moral behaviour, specifically prosocial or altruistic behaviour, ecological behaviour and whether they engaged in behaviour that prized material possessions over altruism. The responses were then scored from high to low. It is in virtue of the score in this part of the study that Frimer and Walker are able to determine whether moral integration is present in a participant. In other words, the predictive validity of autobiographical narratives to identify moral integration is dependent on whether the participants score highly on the moral functioning tests. The moral functioning test is a method independent from autobiographical narratives because it doesn't depend on the narratives of the participants to set a score. For this reason, autobiographical narratives do not identify moral integration.

Verdejo, Victor M. (Pompeu Fabra University)

The Language of Thought, Re-Emerged and Re-Levelled

The Language of Thought hypothesis (LoT) is one of the most well-known and contentious proposals in cognitive science. Echoing the impetus of the Fodorian approach (Fodor 1975, 2008), Quilty-Dunn et al. (2022) have recently offered a revived vindication of the hypothesis as one that is, if not the only, at least "the best game in town". Their insightful paper develops a rich and thought-provoking case by drawing on results of current empirical research across computational, developmental, perceptual, comparative and social psychology.

Their main argument starts with the identification of six distinctive properties of representational format characteristic of LoT structures (Quilty-Dunn et al. (2022, sect. 2): (i) discrete constituents corresponding to individuals and their features; (ii) syntactic role-filler independence; (iii) predicate-argument structure apt for truth-evaluation; (iv) the use of logical operators; (v) automatic inferential promiscuity; and (vi) the representation of abstract conceptual content. They then go on to argue that these properties are present in the kind of representation posited in a number of research areas, including computation and perception studies, and various forms of infant and adult cognition paradigms.

The consideration of LoT in the terms proposed by these authors is surely well-grounded, rewarding and refreshing. Yet their approach would seem to raise more questions than it actually settles regarding our understanding of explanatory aims across cognitive science and the correct characterization of LoT as a hypothesis that is both plausible and open to empirical (dis)confirmation. In this paper, I will attempt to address these issues by providing a much clearer delineation of the levels at which LoT -and indeed rival explanations- are supposed to be operating. In particular, I will argue that these hypotheses are better seen as affording integrated multi-level explanations of cognitive phenomena in a given domain.

One surprising fact about Quilty-Dunn et al.'s target paper is precisely that they pay little to no attention to the proper level or levels of explanation at which LoT is to be formulated (one exception to this is found in fn. 5's remark on computational Bayesian models). But without an accurate statement of this aspect it is hard to get a grip on the relevant sense of representational format operative in their account and the credit that the putative "re-emergence of LoT" deserves.

There are various ways of understanding and identifying explanatory levels in cognitive research and their relation to one another (Potochnik & de Oliveira 2020). There is also room for various positions regarding the

link between computational or psychological levels and mechanistic levels (Boone & Piccinini 2016, Harbecke 2021). For present purposes, we may rely on David Marr's (1982, chap. 1; cf. Newell 1986, Pylyshyn 1984) classic tripartite analysis in terms of function, algorithm and implementation structures (see also Bechtel & Shagrir 2015). Importantly, the consideration of these levels is as such not to be equated with a particular view on heuristics (methodological priority of one level over the others), reductionism (of one level to the others) or the claim that each level, however exactly delineated, is independent of the other two.

One key point is that LoT is certainly undermined when considered to be a hypothesis confined to one of the levels. For instance, one would not think that the functional level is characteristic of LoT: rival accounts would share explanatory targets in terms of what function is computed, either intentionally or extensionally conceived. Exclusive consideration of the level of the algorithm or representation, although perhaps more relevant from the point of view of the format of the representation, is also inaccurate because different algorithmic proposals are, in fact, compatible with a LoT structure and also compatible with various target functions. Most clearly, restricted focus on the implementation level would misleadingly suggest too direct a correspondence between LoT representational formats and neuronal correlates. Nonetheless, while it is true that statements of LoT explanations often omit the implementational level, this does not mean that it is irrelevant. The view that the neuronal level is fundamental or fundamentally constrains explanation in cognitive science is perfectly compatible with LoT so long as the distinction between levels is kept in mind.

By contrast, LoT becomes most plausible and worth-considering when taken to yield integrated multi-level explanations, i.e. explanations based on LoT-structured representations (algorithmic level), whose operations match performance (functional level) and are psychologically real (implementation level). Discussion of the multi-level nature of LoT crucially complements the characterisation of the hypothesis in terms of the six core properties given by Quilty-Dunn et al. Yet the consideration of levels of explanation is also relevant for the proper assessment of the empirical evidence in favour of and against LoT in the different areas revealingly brought out in the target paper and elsewhere. It is especially pressing when comparing LoT to other competitors. I shall focus on Deep Neural Networks (DNN) whose impact and relevance vis-a-vis LoT architectures is hard to overemphasize (LeCun, Bengio, & Hinton 2015). All the same, and in a way that resonates with the dialectics surrounding LoT back in the 1980s, depending on the levels at which DNNs are supposed to operate, these approaches could either just implement LoT structures or constitute genuine multi-level explanatory rivals whose success would eventually involve rejection of LoT.

Finally, the need for clarification regarding levels of explanation is all the more vivid when we examine the complexities and specificities of the admittedly large number of domains in which LoT and other cognitive hypotheses have application. It is unlikely that domains so different as (non-)human perception, implicit attitudes in reasoning or the various forms of Bayesian computation would require anything like the same LoT representational format or do so to the same extent. Proponents of LoT often emphasize that the hypothesis cannot be the whole story about cognition in ways that lead to the endorsement of pluralistic formulations and acknowledgement of a number of LoTs (Mandelbaum et. al 2022). This exacerbates the need to better demarcate the multi-level explanatory contribution of LoT and non-LoT models as well as the rival LoT alternatives in each given domain.

Wang, Yuxing (Massachusetts Institute of Technology)

Hopping through the conceptual space: Diffusely attending to memory in creative thinking

Recent philosophical literature on creativity has pointed to the role that attention plays in creative cognition, and has primarily focused on the role of broadening one's attention to a variety of content in creative thinking. I argue that while these theories capture one aspect of creative thinking, they do not in themselves distinguish the underlying attention mechanisms of creative thinking from that of other related forms of spontaneous thoughts (Kristoff et al. 2016), such as mind wandering. This is likely due to a bias in the literature of over-focusing on the early stage of creativity. I propose an alternative theory that avoids such bias.

Recently in philosophy, Chandra Sripada and Sarah Aronowitz have both considered using the explore/exploit framework to model open-ended thinking (Sripada 2018, Aronowitz 2021). According to this framework, an epistemic inquiry for answering an open-ended question would require a balance between exploring different options for further inquiry, and focusing on one line of inquiry. My proposal applies the explore/exploit model to creative thinking. Creative thinking, such as the open-ended problem of designing a sofa, requires first exploring different ideas, and later, exploiting the most promising idea and further developing and polishing it.

This proposal fits with multi-stage models in psychology and philosophy of creative thinking. In the seminal "geneplore" model proposed by Ronald Finke and co-authors, when thinking creatively the agent first generates some preliminary ideas in the "generate" stage, and then in the "explore" stage, the agent picks and chooses from the generated ideas and further develop some of them into more mature solutions to the creative problem (Finke et al. 1992). The "generate" stage likely requires more diffused attention over a wide range of content, while the second stage requires focused attention on a certain idea.

Previous studies tend to bias towards focusing on the exploration stage, but fail to consider the ability to aptly switch from the exploration to exploitation stage is also crucial for creative thinking. Moreover, I will argue that flexible attention, i.e. the ability to flexibly switch between more diffuse and more focused content, plays a more important role in creative thinking than previously considered, and especially in the process of switching between exploration and exploitation.

I show that recent accounts of the attention mechanisms of creativity overfocus on the exploration stage of creative thinking. One of such theories is Peter Carruthers' account of online and offline creativity thinking (Carruthers 2020). The former requires top-down control mechanisms to inhibit prepotent ideas to allow later, more creative ideas to emerge, and the later requires "leaky attention", i.e. expanding the focus of one's attention to information that are seemingly task-unrelated and later retrieve them to solve the creative task. Both his accounts of online and offline creativity explain the mechanism that expands one's attention in the exploration stage and avoids exploiting ideas prematurely, but does not explain how the agent switches to exploration later on.

Another of such theories is Dustin Stokes' account of incubation in creative thinking (Stokes 2007). Stokes argues that diffusing one's attention prevents someone from fixating on a narrow set of information that might turn out to be unhelpful for completing the creative task. I argue that diffusing attention during the incubation stage is a plausible candidate for the attention mechanisms of the exploration stage but again fails to explain how the agent later switches to exploitation.

To better explain the attention mechanism that underlies the switching from exploration to exploitation, I draw on psychology studies that show that creative people perform better at flexibly switching between global and local perpetual attention (Zabelina, Saporta, & Beeman, 2016). People who are ranked as more creative in solving word puzzle tasks are better in switching between recognizing big letters made up of smaller letters, and recognizing the small local letters that comprise the big ones. This suggests that creative thinking likely requires subjects to switch between focused and diffuse attention rapidly according to task demand.

Given that there is a growing literature on the role of flexible control in skilled action (Christensen 2016, Pacherie & Mylopoulos 2021), and the plausible connection between creativity on the one hand, and skill and expertise on the other hand, it will be interesting to explore other ways that flexible attention plays its role in creativity.

Weger, Daniel (Goethe University Frankfurt)

Structuralist representationalism about phenomenal consciousness

The hard problem of consciousness is the problem of explaining how and why physical states and processes give rise to conscious experience. The core idea of representationalism about phenomenal consciousness is that the phenomenal character of experience is determined by its representational content. In its strongest version, representationalism holds that phenomenal character just is representational content that satisfies certain further constraints. Traditionally, tracking representationalism, as advocated

by Tye (1995, 2000), Dretske (1995), and Lycan (1996), has been the predominant view. It combines the representationalist idea with the tracking theory of mental representation, which claims that a mental state represents, for example, that something is red if and only if it has the function of indicating that there is something red or causally correlates with red things under optimal conditions. However, tracking representationalism faces serious objections: First, phenomenal character as understood by tracking representationalism does neither qualitatively nor structurally match the phenomenal character as manifest in conscious experience (Mendelovici 2018, Pautz 2006). Second, it is committed to color physicalism, which is subject to many objections like the problem of metamerism (Lycan 2001). Third, it cannot account for systematic misrepresentation, which is a genuine possibility (Mendelovici 2013). Fourth, empirical findings suggest that phenomenal internalism is true, i.e., the view that microphysical duplicates necessarily share phenomenal properties (Pautz 2013, 2020). Yet, tracking representationalism must advocate phenomenal externalism because the content attributed by the tracking theory depends on the environment and the history of the subject of experience.

The shortcomings of tracking representationalism have led many philosophers to abandon the idea of representationalism altogether. Moreover, they have given rise to the phenomenal intentionality theory, which reverses the direction of explanation and claims that phenomenal consciousness is the source of intentionality (Farkas 2008, Horgan and Tienson 2002, Kriegel 2011, Loar 2003, Mendelovici 2018, Pitt 2004, Siewert 1998). I will argue that representationalism is generally on the right track, but needs to be equipped with a different approach to mental representation. More precisely, the proponent of representationalism should adopt an internalist account of mental representation because all the major challenges for tracking representationalism are due to its externalist commitments that come with endorsing the tracking theory. While there have been attempts to establish internalist representationalism, they are either non-reductive in spirit and therefore provide no answer to the hard problem of consciousness (Chalmers 2004, 2006, Pautz 2010, 2020) or proceed from tenuous assumptions, such as that there is a language of thought (Rey 1998).

I suggest that representationalists should adopt a structuralist notion of representation that is also found in recent neuroscientific theorizing, such as the Bayesian-inspired framework of predictive processing (Hohwy 2013, Gładziejewski 2016). According to this approach, mental states represent in virtue of their bearing structural similarity to what they represent, just as cartographic maps represent by preserving structural features of what they correspond to (Cummins 1996, O'Brien and Opie 2004, Shea 2014). Moreover, it is characteristic of structural representations that they may be used for guiding the subject's behavior, can be used offline, and allow for the detection of representational error. On this view, then, the phenomenal character of experience can be explained in terms of its representational content, which is based on structural similarity. While this may seem counterintuitive at first, it becomes evident upon closer examination that the structuralist approach can account for the discriminatory abilities involved in conscious experience. Hence, I will argue that this structuralist version of representationalism preserves the idea of representationalism and offers a new twist, while avoiding the pitfalls of tracking representationalism, for it is internalist in spirit.

References

Chalmers, D.J. (2004) 'The Representational Character of Experience', in B. Leiter (ed.) *The future of philosophy*, 153–81. New York: Oxford University Press; Clarendon Press. —— (2006) 'Perception and the Fall from Eden', in T. S. Gendler and J. Hawthorne (eds.) *Perceptual experience*, 49–125. Oxford: Clarendon Press.

Cummins, R. (1996) *Representations, targets, and attitudes*. Cambridge, Mass: MIT Press. Dretske, F.I. (1995) *Naturalizing the Mind*. Cambridge, Mass.: MIT Press. Farkas, K. (2008) 'Phenomenal Intentionality Without Compromise', *Monist*, 91/2: 273–293. Gładziejewski, P. (2016) 'Predictive coding and representationalism', *Synthese*, 193/2: 559– 582. Hohwy, J. (2013) *The Predictive Mind*. Oxford: Oxford University Press. Horgan, T. and Tienson, J. (2002) 'The Intentionality of Phenomenology and the Phenomenology of Intentionality', in D. J. Chalmers (ed.) *Philosophy of mind. Classical and contemporary readings*, 520–33. New York: Oxford University Press. Kriegel, U. (2011) *The Sources of Intentionality*. Oxford: Oxford University Press. Loar, B. (2003) 'Transparent Experience and the Availability of Qualia', in A. Jokic and Q. Smith (eds.) *Consciousness. New Philosophical Perspectives*, 77–96. Oxford: Clarendon Press. Lycan, W.G. (1996) *Consciousness and experience*. Cambridge, Mass.: MIT Press.—— (2001) 'The Case for Phenomenal Externalism', *Noûs*, 35/15: 17–35.

Mendelovici, A. (2013) 'Reliable misrepresentation and tracking theories of mental representation', *Philosophical Studies*, 165/2: 421–443. —— (2018) *The Phenomenal Basis of Intentionality*. New York, NY: Oxford University Press.

O'Brien, G. and Opie, J. (2004) 'Notes Toward a Structuralist Theory of Mental Representation', in H. Clapin, P. J. Staines and P.P. Slezak (eds.) *Representation in mind. New approaches to mental representation*, 1–20. Amsterdam, Heidelberg: Elsevier.

Pautz, A. (2006) 'Can the physicalist explain colour structure in terms of colour experience?', *Australasian Journal of Philosophy*, 84/4: 535–564. —— (2010) 'Do Theories of Consciousness Rest on a Mistake?', *Philosophical Issues*, 20/1: 333–367. —— (2013) 'The Real Trouble for Phenomenal Externalists: New Empirical Evidence for a Brain-Based Theory of Consciousness', in R. Brown (ed.) *Consciousness Inside and Out. Phenomenology, Neuroscience, and the Nature of Experience*, 237–98. Dordrecht: Springer. —— (2020) 'Representationalism about Consciousness', in U. Kriegel (ed.) *Oxford Handbook of the Philosophy of Consciousness*, 405–37. Oxford: Oxford University Press.

Pitt, D. (2004) 'The Phenomenology of Cognition Or What Is It Like to Think That P?', *Philosophy and Phenomenological Research*, 69/1: 1–36.

Rey, G. (1998) 'A Narrow Representationalist Account of Qualitative Experience', *Noûs*, 32/S12: 435–457.

Shea, N. (2014) 'Exploitable Isomorphism and Structural Representation', *Proceedings of the Aristotelian Society*, 114/2.2: 123–144.

Siewert, C. (1998) *The Significance of Consciousness*. Princeton, N.J: Princeton University Press.

Tye, M. (1995) *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, Mass.: MIT Press. —— (2000) *Consciousness, color, and content*. Cambridge, Mass.: MIT Press.

Weksler, Assaf (University of Haifa and the Open University of Israel) and Ori Beck (Ben Gurion University of the Negev)

Weber's Law as a Challenge to Phenomenal Diaphaneity

In this talk, we argue that Weber's Law, a well-established psychophysical finding (Fechner, 1860; Gescheider, 1997; Pardo-Vazquez, et al., 2019; Weber, 1834), poses a challenge to 'phenomenal diaphaneity', the claim that similarities in the phenomenal characters of visual perceptual experiences are due to similarities in the external properties that the experiences make one aware of. Phenomenal diaphaneity is accepted by some representationalists (e.g., Dretske, 1995; Tye, 2009), as well as by some naïve realists (e.g., Allen, 2016; Sethi, 2020).

Weber's Law is presented in terms of just-noticeable-differences (JNDs): If a subject S is presented with a stimulus intensity I, there exists a minimum intensity increase ΔI that provides a 0.75 chance of correctly identifying which of I and I + ΔI is more intense. ΔI is the JND for I (relative to S). Weber's Law states that over a wide range of magnitude intensities I, the ratio between I and its respective JND is constant.

To see how Weber's Law challenges phenomenal diaphaneity, suppose a 22 mm stimulus is the shortest stimulus that S has a 0.75 probability of correctly identifying as longer than a 20 mm stimulus. Then, the JND for 20 mm is 2 mm, and the ratio between the 20 mm stimulus and its JND is 0.1. Weber's Law states that this ratio is constant across a wide range of lengths; and in particular that the shortest stimulus that S has a 0.75 probability of correctly identifying as longer than a 22 mm stimulus is a 24.2 mm stimulus. Thus,

(1) 22 mm are not noticeably different from 24 mm, but 22 mm are noticeably different from 20 mm.

Plausibly, if two stimuli are not noticeably different, then the phenomenal characters of the perceptions of those stimuli are more alike than the phenomenal characters of the perceptions of two stimuli that are noticeably different. So,

(2) The phenomenal characters of perceptions of 22 and 24 mm are more like each other than the phenomenal characters of perceptions of 22 and 20 mm.

According to phenomenal diaphaneity, however, the phenomenal characters of perceptual experiences are fully determined by the external properties the subject is aware of in having those experiences. So, it seems that to explain (2) a supporter of phenomenal diaphaneity has to say that

(3) The magnitudes of 22 and 24 mm are more like each other than the magnitudes of 22 and 20 mm.

The trouble is that (3) seems false — both pairs are only 2 mm different.

We will consider two responses to the challenge. The 'determinables response' suggests that when presented with a stimulus of a given determinate intensity, we only perceive a certain determinable intensity of the stimulus: the determinable intensity encompassing exactly those determinate intensities that the subject could not discriminate (with some fixed probability) from the determinate intensity of the stimulus. In combination with Weber's Law, this idea has two consequences: first, the greater a given stimulus intensity, the more determinable the intensity that the subject perceives. Second, stimulus

intensities which increase by equal amounts will bring about perceptions of determinable intensities whose overlaps increase. To illustrate, if the JND-to-intensity ratio for lengths were 0.1, the determinables response would entail that 20, 22, and 24 mm stimuli bring about perceptions of 18.181-22, 20-24.2, and 21.818-26.4 mm lengths respectively, whose overlaps—2 and 2.38 mm—increase. Because the overlaps increase, perceptions caused by 22 and 24 mm stimuli are more like each other than perceptions caused by 22 and 20 mm stimuli. This could potentially account for (2) and (1) while avoiding (3).

Against this response we argue that subjects perceiving more intense properties undergo perceptions with more determinable phenomenal characters, which contradicts our introspective judgments.

The second response is the 'phenomenal variability response'. This response takes its lead from scientific models, which explain Weber's Law by assuming that the smaller the ratio between two stimulus intensities, the likelier they are to elicit identical (or similar) neural responses in a subject. According to the response, the smaller the ratio between two stimulus intensities, the likelier they are to elicit experiences with identical (or similar) phenomenal properties. For example, stimuli of 22 and 24 mm (whose ratio is ≈1.09) are likelier to elicit experiences as of the length halfway between them (23 mm) than stimuli of 20 and 22 mm (whose ratio is =1.10). This looks like a way of satisfying (2) and (1) while avoiding (3).

However, this response assumes that presentations of length often lead to illusory experiences under ideal conditions. After all, we just assumed that a 24 mm stimulus was likely to elicit an experience as of 23 mm. Because subjects are able to reliably discriminate the stimuli presented to them in Weber's Law tests, and arguably can know on this basis that one stimulus is different from another, it is implausible to attribute rife illusions to them, or so we argue.

Werning, Markus (Ruhr University Bochum) and Kristina Liefke (Ruhr University Bochum)

Remembering dreams: Beyond a representationalist and relationalist analysis of episodic memory

Episodic memories are widely regarded as factive: Linguistic reports of a memory make the presupposition that it refers to an actually existent object and that the properties remembered of the object actually apply to it. Focusing on memories from perceptions – where factivity can indeed be assumed – the two main historical strands in the philosophy of memory, representationalism and relationalism, disagree, amongst others, over (i) whether memory reports should be construed as de re or de dicto, (ii) to what type of entity agents are directed when they episodically remember, and (iii) how the time gap between experience and remembering is bridged. However, what about memories from dreams or hallucinations, where factivity is violated? Here, we can still distinguish remembering from misremembering and confabulation. We show that the representationalist as well as the relationalist faces insurmountable problems in those cases. We develop a new account according to which reference in episodic memory is parasitic on the reference relation in the primary experience – the host attitude. Accordingly, we replace de re and de dicto analyses of memory reports with a de hospite analysis. Referential parasitism overcomes the problems of representationalism and relationalism. We argue that referential parasitism obviates the need to transmit representational content from experience to remembering, as postulated by the Causal Theory of Memory. Minimal traces without representational content suffice. Trace minimalism (Werning, 2020) paired with referential parasitism (Werning & Liefke, 2023) provides a uniform, non-disjunctivist explanation of remembering and misremembering from veridical and non-veridical experiences.

Wiegmann, Alex (University of Bochum) and Louisa M. Reins (University of Goettingen)

The impact of implicit content on lie and truth judgments: A cross-culture study

Lying is a classic philosophical topic and a concept that is deeply anchored in our social life. For this reason, it does not come as a surprise that philosophers explicitly state that a good definition of lying should capture the ordinary use and understanding of this concept. For instance, Carson (2010: 33) writes:

Lying is a concept used in everyday language, and moral questions about lying arise in people's everyday experience. There are no compelling reasons to revise or reject the ordinary language concept of lying—at least the burden of proof rests with those who would revise or reject it. Therefore, consistency with ordinary language and people's linguistic intuitions about what does and does not count as a lie is a desideratum of any definition of lying.

However, if it would turn out that the concept of lying is strongly dependent on the culture people live in or on the language the speak, empirical approaches to finding a definition of lying might be less promising than they seem at first sight. For this reason, we conducted a cross-cultural study with 3660 participants (183 per condition; 366 per country) from ten countries (USA, UK, South Africa, Spain, Chile, Mexico, Israel, Germany, China, and Japan). Moreover, to apply a strong test of whether the concept of lying might be universal, we chose to employ a kind of deception that is not clearly an instance of lying, namely deceptive implicatures, in our study.

Deceptive implicatures are cases in which the speaker says something that is literally true but carries an implicature that (she believes) is false. In recent years, several empirical studies have been conducted to investigate whether people think that it is possible to lie with deceptive implicatures. While the findings are not clear cut, there is quite strong evidence that at least some cases of deceptive implicatures are considered to be lies (see Wiegmann & Meibauer, 2019, for an overview). Moreover, it has been found that some literally true utterances that convey a false implicature are judged to be false (Or, Ariel, & Peleg, 2017).

In the present study, we employed a 2 (judgement: lie vs falsity, between subjects) * 10 (country; between-subjects) * 10 (vignette; within-subjects) mixed design. Participants were presented with ten cases of deceptive implicatures and they were asked either to judge whether the speaker lied ('lie condition') or whether what she said was true or false ('falsity condition'). Here is one example:

Kevin is shocked when he finds an empty pack of cigarettes in the trash. He confronts his wife, Sally. Sally does not want Kevin to know that she secretly smokes from time to time, so she says, "Evelyn was here yesterday, and she is a smoker." In fact, Evelyn visited Sally yesterday and she is a smoker. The empty pack is not Evelyn's pack but the one Sally smoked.

[lie condition] Did Sally lie to Kevin? (yes/no)

[falsity condition] What Sally said was…(true/false)

The overall rate of lie attributions (across all countries and vignettes) was 57%, while the overall rate of falsity attributions was slightly lower (49%). Within each of the ten countries, there was considerable variation in the rates of lie and falsity attributions across the ten vignettes, with percentages ranging from 15% to 85%. Strikingly, however, the result pattern across the ten vignettes was surprisingly similar for the ten countries. The mean correlation (average of all 45 possible combinations of two countries) for lie attributions was r=.9 and r=.89 for falsity attributions. Moreover, the overall correlation of lie and falsity attribution (across all countries) was also very high with r=.92.

We discuss the findings and argue that the high similarity between the ten countries is good news for empirical approaches to finding a definition of lying.

Wolf, Julia (Ruhr Universität Buchum)

Neither Egocentric nor Altercentric – Perspectives and Self-Other Understanding

Historically the view has dominated that infants are initially egocentric and that the ability to take the perspectives of others is a cognitive achievement which children only reach later in development. Against this, Southgate (2020) has recently argued that even young infants are able to take the perspective of others and that this perspective is even encoded more strongly than their own perspective. This allows for explaining the findings from the implicit false belief task where infants appear to be able to take the perspective of another person even when this conflicts with the child's own perspective on the situation (e.g. Kovács, Téglás, & Endress, 2010), while also explaining why they fail the explicit false belief task later in development (Wellman, Cross, & Watson, 2001). It is only once children develop a self-representation that the self-perspective becomes dominant, leading to egocentric errors. I focus on three elements of Southgate's proposal: a) children are initially altercentric, b) once they develop a self-awareness they become egocentric and c) early forms of perspective taking do not require perspective understanding.

In this paper I will critically evaluate Southgate's proposal. While I agree with c) and the criticism of the assumption that infants must start off being egocentric, I will argue that there is evidence that young children are not predominantly altercentric either. For example, on Southgate's view we should find a U-shaped pattern of development on performance on the false belief task where children initially succeed on perspective taking tasks, start systematically failing them around age two, till they improve again at age 4. However, the evidence for this is limited. While there is evidence of early perspective taking from some of the implicit FBTs, these are often subject to replication concerns, sometimes only being partially replicated with older children (Kulke & Rakoczy, 2018; Kulke, von Duhn, Schneider, & Rakoczy, 2018). Similarly, this explanation cannot account for the findings of simultaneous success and failure respectively in the implicit and explicit versions of the false belief task (Clements & Perner, 1994), which seems to indicate that there is something about the task which influences children's perspective taking abilities. In addition, there are reasons to be cautions about the claim that the strengthened encoding of the self perspective due to increasing self-awareness explains children's egocentric error as children have also been shown to make false belief errors regarding their own (past) false beliefs after developing a self-perspective (Gopnik & Astington, 1988).

In line with Southgate, I argue that at an early stage infants are able to represent both their own perspective and that of others. However, I differ in arguing that children are neither predominantly egocentric nor predominantly altercentric. Instead, which perspective is activated is dependent on the situational context. I highlight some aspects of the false belief task which might predispose the activation of the other perspective over the self perspective building on the work of Rubio-Fernández and Geurts (2013). It should be noted that in claiming this I agree with Southgate that this early representation of perspectives occurs without these being as perspectives and that this does not require perspective understanding.

This leaves open the question of how children move beyond this situation dependent activation of perspectives towards a more robust perspective understanding. Arguably this a change which takes place around the age of 4, where children not only succeed in the explicit false belief task (Wellman et al., 2001), but also begin to succeed in a number of other perspective taking tasks (Perner, Stummer, Sprung, & Doherty, 2002). I argue that what underlies this change is a cognitive reorganisation from object centred representations towards more person centred representations in which a perspective is represented as a perspective. This cognitive reorganisation can be fruitfully modelled using the mental files framework (Perner & Leahy, 2016; Recanati, 2012): children initially are able to generate mental files for different perspectives. However, they represent their own perspective and that of others without representing them as perspectives of anyone. These files are at most causally with the person so, for example, the perspective of Mum on an object might be triggered by the presence of Mum and the object. This changes when children develop a self-representation, which provides the basis for the child to develop an appreciation of a perspective being her own. This self representation is required in order for the child to be able to index the file representing their perspective to themselves. When taking the perspective of others children already have representations of others, but their representation of other people's perspectives needs to be re-structured for the perspective to be explicitly attributed to the other person. This means that

children go from having a file of a perspective which is merely causally associated with the other person to having a file which is indexed to the other person (Figure 1). In other words, what is required is a cognitive reorganisation from object centred representations towards more person centred representations in which a perspective is represented as a perspective. It is this reorganisation which allows for the recognition that there are different perspectives on one and the same thing, something which is required for perspective understanding (Perner et al., 2002). In doing so, I aim to clarify not only how perspective understanding develops, but also clarify the notion of 'indexed files', which plays a central role in mental file accounts.

Young, Benjamin (UNR, HebrewU, U. of Copenhagen)

Smell thyself, but not consciously

Our sense of smell profoundly influences our everyday life even in the absence of subjective awareness. Olfaction mediates our food choices, kinship recognition, social acquaintance selection, and mate choices. Previously, it has been argued that olfaction calls into question the general neuroscientific theories of consciousness and the relationship between access and phenomenal consciousness, such that it is best to reconceptualize these types of consciousness as qualitative consciousness and awareness. Extending this framework the talk explores the role that smell plays in allowing us to recognize our embodied material composition and what we can perceive about others from their smell both with and without subjective awareness. The tentative conclusion is that our sense of self is partially constructed through unconscious olfactory processes that track our own smell, kinship relations, and the scent of the other.

Zięba, Paweł (Jagiellonian University)

A Naïve Realist Theory of Consciousness

1. Although naïve realism about perception (a.k.a. relationalism) attributes crucial explanatory roles to perceptual consciousness (see e.g. Campbell, 2002; Travis, 2007), the champions of this view don't say much about the nature and functioning of consciousness. On the contrary, they tend to think that the relation of conscious acquaintance with the perceived item is unanalysable (Brewer, 2011). I believe that this approach to consciousness isn't mandatory for the naïve realist. In this talk, I will argue that naïve realism can and should be complemented with a functionalist theory of consciousness.

2. At bottom, my proposal is a mix of two ingredients: Pure Relationalism about perception (PR) (Stoneham, 2008), and Quotational Higher-Order Thought theory of consciousness (QHOT) (Coleman, 2015). While the original formulations of PR and QHOT are inconsistent with each other, I will show that (i) a small tweak here and there suffices to render them consistent, and that (ii) the proponents of each account have some good reasons to embrace the other, and therefore to accept such small changes in their positions.

3. Naïve realism, as I understand it, encompasses two claims: (i) the phenomenal character of genuine instances of perception (as opposed to hallucination) is at least partially constituted by the perceived items (i.e. mind-independent objects in the environment); (ii) perception doesn't consist in representing the perceived items as being a certain way.

4. PR is a specific version of naïve realism. On this view, perception isn't a state of the subject, but a relation in the world that enables the subject to form certain beliefs and behave in certain ways. Perceiving is having something before the mind, in the sense that the perceived items are available for reason and judgement, but their being perceived doesn't itself involve any mental act. It follows that perceptual phenomenal character is diaphanous, i.e. entirely constituted by the perceived items (cf. Stoneham, 2008, p. 315).

5. If the phenomenal character of perception is diaphanous, it's also consciousness-independent. For diaphaneity entails that whatever has the property of being a constituent of the phenomenal character of a perception, it has that property contingently (it isn't essentially a property of a mental state). Consequently, diaphaneity makes it logically possible that perceptual phenomenal character isn't inherently conscious. The PR-theorist can maintain that perceptual phenomenal character is only conscious when the subject becomes conscious of what they are perceptually related to. Therefore, PR predicts that perception can occur unconsciously.

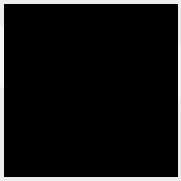
6. Since the possibility of unconscious perception is a well-established hypothesis (see e.g. Kouider & Faivre, 2017; LeDoux et al., 2020), the fact that PR predicts it constitutes a good reason for the naïve realist to specify their view in terms of PR, and thereby depart from the naïve realist orthodoxy of explaining perception in terms of its distinct conscious phenomenal character. However, PR doesn't explain how the subject becomes conscious of the perceived items. PR explains unconscious perception, but not conscious perception. To explain the latter, it has to be supplemented with a suitable theory of consciousness.

7. I will argue that QHOT can play this role. QHOT has been originally formulated by Coleman (Coleman, 2015, 2017) in the context of his panqualityism, a form of Russellian monism according to which the phenomenal character of conscious mental states is grounded in unexperienced qualities, which are fundamental intrinsic properties (i.e. quiddities) of microphysical entities. QHOT explains how the subject becomes consciously aware of such qualities, so that they can determine what it's like to undergo conscious experiences. In short, the idea is that an unconscious sensory state (e.g. an unconscious perception) gets embedded in (or 'quoted' by) a higher-order thought. The upshot is a complex conscious mental state (e.g. a conscious perception), comprising both the sensory state and the higher-order thought that embeds it. The phenomenal character of this complex state is exhaustively determined by the qualities involved in the sensory state.

8. What makes QHOT a perfect candidate for a naïve realist theory of consciousness is that the embedding (or 'quoting') in question is a direct and non-representational relation. Indeed, Coleman (Coleman, 2019) has used QHOT as a basis for a naturalist conception of conscious acquaintance. Even though Coleman's formulation of QHOT is incompatible with naïve realism due to its internalist conception of mentality (Coleman, 2019, p. 70), the embedding mechanism itself doesn't depend on the internalist aspects of Coleman's view (nor does it depend on panqualityism). Hence there aren't any clear obstacles to PR+QHOT fusion. On the resulting view, the embedded sensory state is perception as construed by PR. In effect, the qualities determining the phenomenal character of our conscious perception qua complex state are mind-independent features of the perceived scene. This is the naïve realist theory of consciousness in a nutshell.

9. Mihálik (Mihálik, 2022) has objected that QHOT fails to explain how the subject becomes aware of the qualities that constitute the phenomenal character of their conscious experience. The objection rests on a conceivability argument analogous to the well-known zombie argument against physicalism (Chalmers, 1996). Let A-zombie be my exact replica with respect to both the qualities I am aware of and the structural properties constituting my QHOT mechanism, who isn't aware in the relevant way of any quality. Mihálik argues that: (i) A-zombie is conceivable; (ii) A-zombie is possible; (iii) QHOT is false because it's possible that QHOT's conditions for consciousness are met and the relevant experience doesn't occur anyway. According to Mihálik (Mihálik, 2022, p. 1437), the panqualityist can't dispute the move from (i) to (ii), because they need to make an analogous move to justify their rejection of physicalism. However, the PR+QHOT-theorist not only doesn't have to accept that move, but also has independent motivation to reject it (Zięba, 2022, pp. 21–24). Besides, PR is neutral regarding the panqualityism vs. physicalism debate, so one can endorse PR+QHOT and drop panqualityism. This constitutes a substantial reason for the QHOT-theorist to endorse PR.

POSTERS



Bystranowski, Piotr (Jagiellonian University)

Normative ignorance and the folk concept of law

"Sorry, I didn't know I wasn't allowed to do that." This is a frequently heard and often valid, or even convincing, excuse. In the context of many social rules, such as moral norms, local customs, or rules of etiquette, an ignorant transgressor has good chances of being forgiven and merely informed about the broken rule and reminded to abide by it in the future.

Legal rules are different, though. If there is any legal principle about which lay people and professional lawyers profess a comparable level of confidence, it is the ancient adage that everybody is presumed to know the law and that ignorance of law offers no excuse. While legal and moral philosophers continue to be puzzled by how an individual can be legally sanctioned for failing to follow a rule they did not even know existed, the harsh principle generally remains in place across legal systems. Unlike philosophers, regular people appear mostly comfortable with this distinctively legal way of approaching normative ignorance.

Assuming such an asymmetry between legal and non-legal ignorance, this project asks whether something interesting about the folk concept of legality can be uncovered by studying how people react to actors ignorantly violating different kinds of rules. Are there some factors that make people both believe a given rule is legal and not excuse a person who violated that rule out of ignorance?

I will present the results of an exploratory correlational study in which I confront participants with a battery of social rules, ranging from statutory and case law provisions through customs of informal social groups and household rules. While one group of participants is asked about the degree to which a given rules is law-like, the other group decides on the responsibility of an actor who violated such rules out of ignorance. The results paint a complex yet fascinating picture, in which both the ascription of legality and non-excusable ignorance are independently determined by multiple situational factors. Rich textual data produced by participants also point clearly to tensions embedded in the way people delineate the legal domain.

Fogd, Dóra, Ernő Téglás, and Ágnes M. Kovács (Department of Cognitice Science Central European University)

Spontaneous tracking of other agents' inferences

Successful navigation in the social world requires taking into account what other agents' believe or know. Importantly, others' actions may rely not only on beliefs formulated based on what they witnessed, but also what they could infer from the available information, via deduction. Representing what conclusions others may draw, from the beliefs they hold, may largely extend the scope of actions people can prepare for and thereby contribute to the flexible adjustment to others' behaviour. Despite its arguable importance, little is known how human adults track the logical inferences of others, specifically whether it takes place spontaneously, just like the tracking of other's false beliefs does according to multiple studies (see e.g. Kampis & Southgate, 2020).

To address this issue, we conducted four online experiments, in which adults were presented with picture sequences in which animals invisibly hid in three boxes, while an agent was present. After the content of two of the boxes was revealed, participants could infer the identity or the location of the hidden animal by applying disjunctive syllogism ('not A, not B, therefore C', Experiment 1-2) or combining disjunctive and conditional reasoning (Experiment 3-4). At the end of the trials, they had to rate how likely it is that a target animal hid at a certain location, either from their own (SELF trials, starting with YOU prompt) or from the agent's perspective (OTHER trials, starting with SHE prompt). Crucially, the conclusion the other agent could draw either corresponded the participants' own (true belief trials) or she ended up representing two disjunctive alternatives (underspecified belief trials), as a result of witnessing only one box-opening (Experiment 1-3) or holding an incorrect prior belief (about a certain animal's habitual hiding location;

Experiment 4). We tested whether participants take into account that the other agent represents two alternatives and rate both as likely options on the OTHER underspecified belief trials; and most importantly, whether they represent both alternatives also when this is unnecessary, resulting in an altercentric bias in their own ratings for the alternative the other considers 'possible' on the SELF underspecified belief trials.

Experiment 1 (N=35) revealed similar ratings of the animal's actual hiding location and the alternative only the agent considered possible, on those underspecified belief trials (p= .159), where participants had to perform the judgements from the other's perspective. Crucially, ratings were higher for the 'possible' than the impossible alternative also on the SELF underspecified belief trials (reflecting an altercentric bias), indicating spontaneous representation of both alternatives (p= .014). Experiment 2 (N=34) replicated these findings, showing that adults spontaneously track others' inferences not only about object location but also about object identity. The results of Experiment 3 (N=34) and 4 (N=35) indicated that adults may be able to spontaneously track even multiple, consecutive inferences of others, where task-relevant attributions require the combination of two logical rules, but only when the agent's conclusions are derived from her stable beliefs and appropriate attributions do not require constant monitoring of what information the other has access to.

Gorman, Lenka (Ludwig Maximilian University of Munich)

Choice Enhances Touch Pleasantness

The power of choice in influencing our preferences and behaviour is well-documented in the field of psychology. When we have the ability to choose between options, we tend to value the option we choose more than one that is imposed on us (e.g.: see Leotti et al., 2010 for overview). This phenomenon, known as choice-induced preferences, has been studied extensively in the context of extrinsic rewards such as money or material goods. However, the impact of choice on intrinsic rewards, such as pleasant sensations, has received less attention.

Using affective touch as a test case, we investigate whether choice could increase its intrinsic rewards. Affective touch is a type of touch that is experienced as pleasant or emotionally meaningful and can have a profound impact on a person's well-being (e.g.: McGlone et al., 2014, Field., 2010). We hypothesized that giving participants a choice over certain aspects of a tactile stimulus administered by an experimenter would make the touch more pleasurable, thereby increasing its intrinsic rewarding value.

To test this hypothesis, we manipulated the kind of choice participants (N=25) had over two different types of touch stimuli (fast or slow). The first kind was of low relevance and involved choosing the colour of the glove worn by the person administering the touch. The second kind of choice was of high relevance and involved choosing the location on the participant's arm where they would be stroked. For each kind, participants would either be offered to choose themselves (choice condition), or the computer would randomly choose for them (no-choice condition). We used both direct methods, by asking participants to rate the pleasantness of the touch they received, and indirect methods (pupillometry) to measure changes in pupil size as an indicator of arousal.

The poster will present the detailed results of the study, which supported our hypothesis that choice can increase the intrinsic rewarding value of affective touch. Participants who had a choice over the location of the touch on their arm reported a higher level of pleasantness than those who did not have a choice. Additionally, participants who had a choice reported a higher level of arousal in anticipation of the touch, suggesting that choice can increase the level of excitement and anticipation associated with touch.

Interestingly, the relevance of choice did not significantly impact the participants' ratings of pleasantness. This suggests that allowing individuals to choose, regardless of how relevant the choice is to the actual touch stimulus, can positively enhance their perception of the physical contact they receive.

In conclusion, this study provides insight into the power of choice and consent in enhancing the intrinsic rewarding value of affective touch. This has implications for how we understand the role of choice in shaping our preferences and behaviours, particularly in the context of interpersonal relationships and therapy settings.

References:
Field, T. (2010). Touch for socioemotional and physical well-being: A review. Developmental Review, 30(4), 367–383. doi:10.1016/j.dr.2011.01.001
Leotti, L. A., Iyengar, S. S., and Ochsner, K. N. (2010). Born to choose: the origins and value of the need for control. Trends Cogn. Sci. 14, 457–463. doi: 10.1016/j.tics.2010.08.001
McGlone, F., Wessberg, J., & Olausson, H. (2014). Discriminative and affective touch: sensing and feeling. Neuron, 82(4), 737-755. doi.org/10.1016/j.neuron.2014.05.001

Kispál, Anna (Central European University) and Ágnes Melinda Kovács (Central European University)

Knowledge vs. Belief: Exploring The Underlying System(s)

A recent proposal in Theory of Mind research suggests that factive and nonfactive mentalization should be separated and studied as equally important parts of Theory of Mind. Factive mental states, such as knowledge and ignorance, are tied to the way one takes the world to be. On the other hand, nonfactive mental states, including false and true beliefs are not tied to the way one takes the world to be (Phillips et al, 2021). While a great body of research explains how false belief and knowledge attribution happen in various populations, there is still not enough empirical evidence about how different mental state attributions relate to each other.

The current study aims to explore whether factive and nonfactive mental state attributions recruit the same representational structures, thus belong to the same or separate systems. The relationship between factive and nonfactive mentalization is studied by examining the way updates happen between them, relying on the assumption that transfer within one system should occur faster than between systems. That is, if factive and nonfactive mentalization belong to the same system, updating a nonfactive mental state to another nonfactive mental state (e.g. true belief to false belief), should be easier, than updating from a nonfactive mental state to a factive mental state (e.g. true belief to ignorance).

In the experiment adult participants watch animated videos of a ball moving and hiding between two boxes, while an Agent is either witnessing or not witnessing the movement of the ball. During test trials, updating a previously attributed mental state is needed. For instance, the Agent witnesses the ball moving into a box, then an occluder covers her view and the ball is changing its location, causing her to have a false belief about the location of the ball (false belief attribution). Afterwards, the occluder lowers and the ball changes its location again, but now the Agent is knowledgeable about it (update from false belief to knowledge). Various versions of such mental state updates are shown to the participants (from false belief to knowledge/true belief/ignorance and from true belief to knowledge/false belief/ignorance). At the end of each video, both of the boxes open at the same time and the ball reappears at the location that is congruent with reality. The reactive saccadic reaction to the outcome of these scenes is measured, as a proxy for the speed of updating from certain representations or systems to others (Kenward et al., 2017).

Since data collection is ongoing, final results of the experiment will be presented at the conference. However, pilot data from eight participants suggests that updating from false belief to true belief (M = 969 ms) and from true belief to false belief (M = 972 ms) resulted in faster saccadic reaction times than updating from false belief to ignorance (M = 1024 ms) or knowledge (M = 1007 ms). This shows a faster update within nonfactive mental states and suggests the possible separation of factive and nonfactive mentalization into two systems.

Li, Junyu (Max Planck Institute for Evolutionary Anthropology), Manuel Bohn (Max Planck Institute for Evolutionary Anthropology) and Daniel Haun (Max Planck Institute for Evolutionary Anthropology)

Perception of time intervals and temporal planning in great apes

Time perception is common in human conscious experience, and behavioral data has indicated its presence in some animals, including non-human apes. On one hand for time sensing, operant conditioning literature on various animals including great apes show subjects learn to anticipate fixed-time intervals after extensive training; on the other hand for future-oriented planning, research on wild chimpanzees indicates that they plan foraging journeys with time awareness, and behavioral experiments on apes in captivity show evidence of mental time travel and episodic-like memory. However, how accurate great apes can estimate periods of time and how they integrate the estimates with future planning are not understood, hindering further investigation on whether apes have metacognitive awareness of time perception.

In this study, we investigate chimpanzees' estimation of time intervals with a fixed-interval anticipation task and cognitive modelling. We examine a chimpanzee group's response to a novel reward appearing at a regular interval and use the animals' approaching movement to infer their estimation of time. We place a juice fountain among a chimpanzee group (N=19) in its home enclosure at Anonymous Zoo. The fountain ejects juice for 20 seconds every 15 minutes (5 cycles x 15 days); the juice's transient availability makes anticipatory approaches much more rewarding than reactive approaches. We record overview videos and quantify the animals' movement by detecting their locations with a deep learning model (custom-trained MaskRCNN model). The group's response to the reward is described by a time series of the group's mean distance to the fountain. With agent-based modelling, three single-strategy models are simulated: only-reactionary, only-anticipatory and only-indifferent, from which mixed models are created, representing various compositions of the three strategies in the group. The real group strategy mix will be inferred by fitting the experimental time series. Within the agents using anticipatory strategy, we infer cognitively relevant parameters including the agent's estimate of next reward and the agent's motivation for obtaining the reward.

Data collection is undergoing and is planned to finish by May 2023. Exploratory analyses will aim to show the prevalence of anticipation in the group, the group's estimation of reward interval, and how the group strategy composition changes over more exposure to the reward. Additionally, animals' locations between rewards will be analyzed to discuss whether an animal's self-placement in anticipation of a reward indicates its judgement of the timely reachability of a location and self-awareness of speed. For comparison between the experimental condition and the normal daily space use, the group's location data over 3 months before the experiment is used as baseline.

Montessori, Auke (Washington University St Louis)

The Mixed View and Multisensory Experience

It seems clear that, introspectively, the senses feel quite different. When we see something, it feels directly present to us. However, smelling something feels indirect. We pick up a smell, but usually don't feel like we have a direct connection to the sensed object itself. Despite such differences, the overwhelming majority of philosophers takes a unified approach to the metaphysical nature of the senses. That is, all senses have the same nature. They are all forms of naïve realist direct awareness, or all intentionalist contentful mental states or et cetera. A few philosophers have suggested that this approach is misguided. Instead, we should explore a mixed view on which the various senses have different natures. Perhaps vision consists of direct awareness, whereas smell is a contentful mental state. This would explain the at times striking phenomenal differences between the senses. However, to my knowledge, nobody has investigated the mixed view in detail. This paper aims to do so. In particular, I discuss a major problem for the mixed view,

namely multisensory perception. Note that by 'perception', I will mostly mean perceptual experience rather than subpersonal processes.

The mixed view is attractive to those who think that phenomenology should be kingmaker in discussions about the nature of perception. The senses feel different in certain ways, so if phenomenology is kingmaker, then the mixed view is automatically attractive. Further, a mixed view is much more flexible when it comes to responding to problems with specific senses. Puzzles involving vision are different from puzzles involving hearing, touch or smell. By allowing the senses to have different natures, we can respond to these puzzles more adequately. There is no reason to assume that an commitment that works for vision should also work for audition.

A potential problem for the mixed view is that the senses often closely interact, and sometimes even create single multisensory experiences together. The multisensory nature of perception is difficult to account for if the senses differ metaphysically. How can content and direct awareness interact, let alone form a single experience together?

Two senses with different natures together creating one multisensory experience is relatively easy to account for. The senses work together at the sub-personal level to create a unified multisensory experience, containing elements from both senses. There is a puzzle about the nature of such multisensory experiences, but a potential account is suggested.

Interaction is more difficult. I focus on the interaction between naïve realist and intentionalist senses. Intentionalist experience mostly involves mental content, whereas naïve realist experience mostly involves direct awareness of the environment. I suggest that attention is what allows them to interact. Certain contents can help us focus, making us aware of new things. Likewise, direct awareness can help the intentionalist sense focus on particular features of the environment. More needs to be said, but this sketch gives us an idea of how multisensory interaction amongst metaphysically different senses is possible.

Nickel, Erik Samuel (Universität Osnabrück, Università degli studi di Trento)

Implicit Gender Bias in the German Federal Elections (2021) – Is there an Effect of the Masculine Generic Form?

In the present study, we investigate the state of the masculine generic form in language comprehension in the German language. For this, we employed a self-paced reading task with a vignette introducing the then-upcoming German federal election in 2021.

Background

The German language is a fully grammatically gendered language, compared to other languages which are completely genderless . Hence, all nouns are assigned one of three grammatical genera: feminine, masculine or neuter. For our study we considered role nouns which refer to a specific social or occupational role of a person . Since German genera are closely connected to the gender and sex of a person, there is a debate how to use the different gender forms in German and, specifically, how to produce a generic form. The masculine generic form is regarded by many to be the standard for such a generic form. Past research made evident that this approach is unsatisfactory, in the regard, that it does not capture the intention of the speaker of producing either no or evenly spread gender information. To the recipient, the male gender is vastly over-represented . Previous studies have shown that in German, political role nouns are predominantly male represented, which is support by the mostly male staffed parliament . Contrary to these past overall trends, the position of the German chancellor before the elections in 2021, was held by Angela Merkel, for sixteen consecutive years. This means in turn that for a whole generation growing up in Germany, the highest political position was held by a woman. To incorporate all these details into one coherent study, a design by was adopted, with some overlap in the adaptation to the followup study by . This study examined the status of the masculine form in German. We hypothesized that the masculine generic form would be dominant compared to the other gender forms. This would mean that the widespread use of the generic masculine form would have a stronger effect on processing preferences than the recent stereotype of a female chancellor.

Method

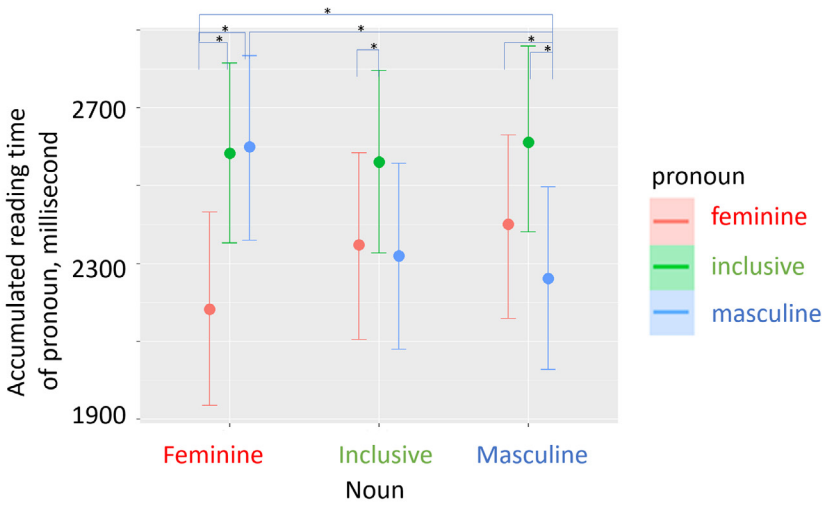
For this purpose, an experiment was conducted preceding the German federal elections in September 2021, recruiting n = 56 (female: 29) university students as participants. This study utilized the method of self-paced reading and included a measure for the expectation for the gender of the next German chancellor. The reading times of masculine, feminine and inclusive forms were examined to deepen the understanding about their roles in the German language. As a vignette, we used a non-priming stage-setting sentence, determined in a pilot study. The actual stimulus sentence comprised two parts. The first mentioned the noun "chancellor", the second used a personal pronoun referring to the chancellor. The noun and the pronoun each varied in their genus. For each there were the three options of masculine form, feminine form and inclusive form. The masculine and feminine forms are the standard forms "der Bundeskanzler" and "die Bundeskanzlerin" respectively. Since there is no standard for an inclusive form we picked one of the most prevalent forms, which is the masculine root of the noun with the feminine suffix separated by an asterisk, "der*die Bundeskanzler*in" (the asterisk is commonly read as a glottal stop). Each participant read each of the nine possible combination of noun and pronoun in a standard moving window self-paced reading paradigm. Before, their belief of gender of the chancellor was estimated in a separate task.

Results and Discussion

The analysis was conducted using Bayesian inference methods. There was no effect of the belief in different gender, of the next chancellor, the reading time of the noun. However, there is an effect in the reading times of the pronoun. The interpretation of the results of the reading time of the pronoun is suggestive regarding the way the noun is perceived.

Comparing the congruent pairs, in which noun and pronoun have the same gender, the masculine pair is read significantly faster than the feminine pair and nearly significantly faster than the inclusive pair. The difference between the inclusive and feminine pair is not significant. When comparing each of these congruent pair to all other pairs which have at least one of the congruent constituents(e.g. feminine-feminine pair versus feminine-masculine, feminine-feminine versus feminine-inclusive,...) The masculine-masculine and the feminine-feminine noun-pronoun pairs seem to follow one pattern. They are always read significant or nearly significant faster than the incongruent pairs. Besides the case, of the inclusive noun followed by a masculine or feminine pronoun respectively. It seems, that the inclusive noun form is perceived generic, open for other gender forms. This gets evident by the lack of low level processing difficulties expected from grammatical incongruent forms. This does not hold for the inclusive pronoun, which was a novel form in this study. That the inclusive-inclusive noun-pronoun pair is read significantly faster than the feminine-inclusive is surprising and needs further to investigate and validate any possible stable effect.

Overall, the masculine form is dominant to the other forms since it is read faster than most case, so we can clearly see a male stereotype support by this but not a true masculine generic form. This stereotype can not be grounded by the fact that there was a long term female chancellor, but rather by the use of masculine generics. This is support by the lack of effect for different gender belief of the next chancellor. We got primary insights, into the inclusive noun form, which provides what we expect from a generic form. It does not reduce the reading time significantly, no matter the gender assigned to it by a subsequent pronoun. Furthermore, we see that this inclusive noun form is not adaptable to pronouns. This indicates that the past active use of the noun form is what provided its advantage as a generic form.



Accumulated reading times of the pronoun and the four consecutive words in milliseconds. The asterisks mark the significant interactions with a posterior probability .

Pham, Que Anh (University of Massachusetts, Boston)

Using Theory of Mind in Memory Guided Planning

The problem of other minds, psychologically referred to as Theory of Mind (ToM), can be a component of other cognitive tasks. Research on episodic future thinking (EFT) has suggested that ToM may be involved in children's development of this ability (Atance & Jackson, 2009). Simulation Theory may underlie the convergence of EFT and ToM as both involve modeling a state that is different from one's own. Memory-guided planning (MGP), a subset of EFT where children use past memories to plan for the future, may occur in the context of planning for somebody other than oneself. Inhibitory control (IC) seems to be the executive function most relevant to MGP as children must suppress irrelevant memories for goal-directed planning. Given the potential relationships between EFT, ToM, and IC, and to address the lack of studies examining EFT and ToM in tandem, we devised a novel task to investigate how children deploy MGP when planning for an animal character. In addition, we tested EFT, ToM, and IC independently using separate tasks to study the association between these abilities.

We tested 33 children [18 3-year-olds (M=43.0m, SD=2.87, 13 female), 15 4-year-olds (M=53.2, SD=3.51, 7 female)] over Zoom and plan to test 5-year-olds as well. Each child completed an MGP-ToM (Ice Cream Machine – ICM), IC (Dimensional Change Card Sort – DCCS), ToM (Not-Own-Belief – NOB), and the EFT (Spoon) tasks. In the novel ICM task, children were taught that they can make chocolate/vanilla ice cream in an "ice cream machine" using yellow/red coins. Children were then introduced to two animal characters, each of whom enjoys either chocolate or vanilla ice cream (counterbalanced). In test phase, children were asked to retrieve ice cream for each of the animal characters using one of the two coins. To succeed on this task, children must use ToM to infer which flavor each animal character intends to get from the ice cream machine given their preferences, and subsequently, use memories of which coin will make that flavor to make a correct selection. In addition to overall accuracy, we were interested in whether children would perform better when the animal character's ice cream preference matched their own (congruent) rather than when it did not (incongruent).

Performance on ICM was significantly above chance for 4-year-olds (M=.80, t=3.52, p=.003) but not 3-year-olds (M=.58, t=.97, p=.34). The congruent condition was significantly above chance (M=.70, t=3.03, p=.004) and the incongruent condition was significant at the .05 level but not at the .01 level (M=.67, t=2.35, p=.03). Regression analysis results with NOB, DCCS, and Spoon task as predictors of ICM performance are not currently significant, but data collection is ongoing.

Preliminary results suggest that MGP improves across early childhood, but sources of variability in MGP-ToM performance are still unclear. Children appear to have an easier time planning when taking on another's congruent compared to an incongruent perspective, possibly because it is less challenging for children to model someone else's behavior when their desires match the child's. These results are consistent with Simulation Theory.

Reimer, Ludmila (Ruhr-University Bochum)

Saliency – Resolving Contradictory Findings in Gesture Research

Gesture research is employing a wide range of methods nowadays and gets more and more sophisticated: we have studies employing EEG and fMRI, our video encoding technology got better and also the quality of videos themselves vastly improved in the last few decades. And the more varied this research gets, the more seemingly contraindicatory findings occur. We want to build our case by looking at one example, namely the seemingly contradictory findings between three studies as identified in Arachchige's et al. (2021) review paper and argue that by introducing a gesture's saliency in discourse as defined by several factors, can help resolve these specific contradictions and can be also employed to other studies as well.

We define saliency as a factor that captures a listener's/onlooker's degree of attention towards the (co-speech) gesture. It influenced by the visual context a gesture is performed in, as well as the communicative embedding.

In their section on EEG, more precisely on the N400 component, Arachchige et al. compare results by Kelly et al. (2004) and Wu and Coulson (2005, 2007) to Habets et al. (2011). Wu and Coulson observed an increase of the N400 when a soundless gesture clip and an unrelated word were presented together, even then the probe word appeared 1000ms after the offset of the gesture clip. Kelly et al. presented their gesture 800ms after the onset of the probe word, also observing an increase in the N400. However, Habet et al. found that when a gesture and its corresponding utterance were presented 360ms apart, no increase in the N400 was found. One possible explanation offered by Arachchige et al. is the different nature of the stimuli used: Wu and Coulson presented stimuli that were supposedly less ambiguous than the ones of Habets et al., on the account that the gestures used by Wu and Coulson were required to be explicitly judged regarding their relatedness to gestures and Habets' et al. gestures were hardly understandable without speech. However, both experimental series their stimulus rated prior to the EEG study and both found a consistent and significantly higher match rating for the congruent pairs than for the incongruent ones, so the materials used by either of the researchers were equally (un)ambiguous and had a similar degree of iconicity.

If not for the iconicity or ambiguity, what else could the contradictions? Luckily, both Wu and Coulson and Habets et al. provide still images of their stimulus material, Kelly et al. only provides a description. Looking at the framing of the speaker, Wu and Coulson capture the speaker's body from an angle, standing and visible from the mid-thigh up. The face of the speaker is visible. The actor wears casual, loose clothing, and the color contrast between clothing and body/hands is low. Kelly et al. have their actor sit behind a table with objects in front of him, so he is probably only visible from the waist up. Face and hands are directed toward the camera and in full view. No information on the clothing is provided. Habets et al. have their actor sit on a chair, her head out of frame, but her hands and arms are fully visible. Her body is facing the camera. She is wearing casual clothing, a long-sleeved black shirt, and the color contrast between the body and the hands is high. – The striking difference here is that Habets et al. do not show a head. At first glance, this should suggest that subjects pay more attention to the gestures and draw a stronger connection between gesture and speech, especially given that in the case of Wu and Coulson the probe word was not even played as an audio file, but simply shown on screen – however, just as in the case for the material used by Kelly et al., the subject had visible mouth movements during gesturing. So even though Wu and Coulson's material was "further apart", it could have elicited the association of the speaker uttering the probe word in the silent video before. In the material of Habets et al., the voice-over could stem from another person, potentially being perceived as not coming from the person on screen. To complicate matters, both Kelly et al. and Habets et al., have their actor's bodies face the camera, which could be interpreted as more communicative towards the speaker. Here, the very different presentation of speakers' bodies and probe words open make a direct comparison of how subjects rated the speakers' communicative intents towards them impossible. Only the material used by Kelly et al. provides a clear communicative intent of the speaker towards the subject.

Looking at this partial analysis, we can see that the following factors could influence the saliency: Visible body parts (full body, mid-thigh up, torso, waist up, chest; head in frame; hands confined to visible area); orientation of the body (towards the camera, at an angle, moving between the two); clothing of the subject (e.g., black clothing obscuring the body, but not face and/or hands); visibility of the head (not in the frame, masked before filming, obscured in post-production – area and degree of blurring/masking); parts of the gesture in the clip (prep, hold, stroke, retract); and the arrangement of gesture parts (continuous video or spliced together). We group these factors as the variance in visual presentation of gestures. Further, we will establish how saliency is influenced by speech (speech content, speech component coinciding with the gesture, type of speech-gesture combination), by other movements (of other gestures, presence of meaningless hand movements, presence of actions), and by the experimental task (in)direct assessment of speech and gestures, unrelated distraction task).

This detailed dissection of three experiments seems tedious, but it reveals that even though they are very similar on the surface, they differ significantly regarding their stimulus material. This needs to be taken into account while comparing gesture studies with each other. We want to offer a set of factors that can be used to determine the saliency of gesture stimulus material to enable systematic comparisons.

References

Habets, B., Kita, S., Shao, Z., Özyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. Journal of Cognitive Neuroscience, 23(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Kandana Arachchige, K. G., Simoes Loureiro, I., Blekic, W., Rossignol, M., & Lefebvre, L. (2021). The Role of Iconic Gestures in Speech Comprehension: An Overview of Various Methodologies. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.634074

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. Brain and Language, 89(1), 253–260. https://doi.org/10.1016/S0093-934X(03)00335-3

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. Psychophysiology. https://doi.org/10.1111/j.1469-8986.2005.00356.x

Wu, Y. C., & Coulson, S. (2007). Iconic gestures prime related concepts: An ERP study. Psychonomic Bulletin and Review, 14(1), 57–63. https://doi.org/10.3758/BF03194028

Rivadulla Duró, Andrea (University of Antwerp)

Lost in translation: Iconicity and Representational Silence in emotion

Emotions can be insensitive to certain attributes of a situation: Fear of flying is not always reduced by remembering aircraft probabilities. A large body of evidence shows that information on probabilities, large numerical counts, or intentions is frequently disregarded in the elicitation and regulation of emotions. To date, no existing theory comprehensively accounts for the features that tend to be overlooked by emotion. In this paper I call attention to the common denominator of such features: they cannot be perceived nor contribute to the iconic representation of events. For instance, the exceedingly low probability of a plane crash does not affect its imagistic representation (i.e., the iconic representation of the event is silent about the event's probability). This paper introduces the Iconic Prioritization Hypothesis, positing that the prioritization of the iconic format in emotion can explain the neglect of information that is representationally silent in this format. Delving into the causes of this format prioritization, I argue that emotion may favour iconicity as it is the format of immediate information about our surroundings (perception) and of stored first-hand evidence (episodic memory). Lastly, the hypothesis's compatibility with philosophical theories of emotion causation and its implications for experimental research are examined.

Severs, Liberty (CFCUL, University of Lisbon)

Embodied selfhood and joint agency during HRI

Artificial technologies have an increasing impact on human lives due to their embeddedness within our social and cultural environment (Gallagher 2000; Haggard et al. 2002; Pacherie 2007; Sahaï et. al. 2022; Qin et al. 2020). In order to better understand this impact, it is timely to understand the effect(s) of these technologies in our lived world: not only how we create these systems in our likeness, but also – in turn – the manner in which these systems might change us (Caspar et al., 2015; Tsakiris, 2017; Ciaunica & Crucianelli, 2019). This experimental study leverages a specific kind of artificial agent – a humanoid robot – to understand the relationship between the sense of embodied selfhood and social joint agency (Fussell et. al. 2008; Thellman et al. 2022; Rosenthal-Von Der Pütten et. al. 2014). To this end, we use a task whereby both artificial (i.e. a social robot) and biological (i.e. human) agents interact during a cooperative task performed in dyads of either two humans or a human and humanoid robot (Sahaï et al, 2022). We acquire behavioral, physiological, subjective report, and neural data to disentangle the various effects of these interactions on the agential bodily self.

We aim to determine: 1) whether (and in what manner) artificial agents impact our self-experience and

our interactions within the social environment; 2) the variability of these alterations to self-experience and our interactions with the social environment; and 3) the influence of attitudes toward/beliefs about artificial agents in shaping these alterations to self-experience during social interactions. We will share preliminary results to demonstrate how the agential bodily self may be affected by interacting with artificial agents.

Stawski, Filip (Kazimierz Wielki University in Bydgoszcz)

Affordances and mental disorders

The aim of the poster is to analyze whether the conception of affordances can be used to describe aspects of mental disorders. The category of affordances, coined by James Gibson and applied in cognitive science, especially in situated cognition conception, is supposed to help explain the direct perception and interaction of the agent and the environment. However, the term affordance is understood differently and has been applied to a range of theoretical and practical problems in recent decades.

According to neuroscientist Paul Cisek, action selection in a specific environmental situation is the result of a constant competition of alternative affordances. In this view, affordances have a concrete neurobiological substrate that reflects specific motor reactions and cognitive processes. Effective decision-making is influenced by various sources of inputs assessed in terms of their potential consequences, resembling a competition among potential actions.

In the poster, I will consider whether a wide spectrum of symptoms accompanying mental disorders, such as hallucinations, delusions, and inappropriate decision-making, can be understood as a weakening of adaptive affordances, which consequently lose the competition with alternative strategies. This general idea will be presented as an example of a disorder of decision-making and self-control by people suffering from addiction.

Taylor, Lauren (University of Stirling); Gema Martin-Ordas,* and Eva Rafetseder* (University of Stirling) (*equal contribution)

The Relationship between Episodic Memory and Counterfactual Thinking in Children

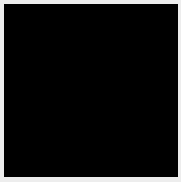
Episodic memory and counterfactual thinking are two cognitive processes that are critical for children's cognitive and social development. Episodic memory is the ability to remember personal past events, while counterfactual thinking is the ability to imagine alternative outcomes to past events. While these have been studied extensively as separate abilities in children, the interaction between the two, termed episodic counterfactual thinking, has so far only been studied in adults, and is less well understood in children. Thus, the aim of the current study was to test children's episodic memory and counterfactual thinking to understand how these abilities interact developmentally.

Since mature reasoning with counterfactuals becomes evident from the age of 6 and episodic memory typically emerges between ages 3 to 5, for the present study we focused on children aged 3 to 8 years (n=64 per year of age). This would allow us to depict a full developmental picture of both cognitive abilities. In the study, children were presented with a hiding-toy task. Specifically, in the most basic version of the task, they were shown three differently coloured boxes and three familiar toys. Children were then asked to put each toy into a different box. Subsequently, the experimenter moved one toy to a different box (e. g., a ball from a yellow box to a pink box), such that one box was now empty, one contained one toy and one contained two toys. Each box was then closed with a lid. After a 3 min delay, children were asked, in

a counterbalanced order, a counterfactual question ("If the ball hadn't been moved to the pink box, what would be in the pink box?") and two episodic memory questions ("Which toy did you put in the pink box in the beginning?", "Which toy is in the pink box now?"). Children were moved to the next level (e.g., hiding 4 toys in 4 different boxes), if their responses to the counterfactual and episodic memory questions were correct. This was done to increase the episodic memory load.

If the two abilities have a bearing on each other, the order in which the test questions were asked – either counterfactual questions or episodic memory questions first – should affect both test questions. We predicted that asking episodic memory questions first would increase children's performance on counterfactual questions, while asking counterfactual questions first would lead to a different performance in episodic memory, although the direction was unclear. Preliminary data indicates that the episodic memory question being asked first results in higher counterfactual thinking scores. However, the difference in episodic memory scores was negligible between the question orders. The results will be discussed in the context of the wider research into episodic counterfactual thinking as well as more specifically in relation to the questioning of child witnesses and how their memory may be affected if asked to reason with counterfactuals.

ESPP SATELLITE WORKSHOP ON OLFACTORY PERCEPTION



Chiara Brozzo (University of Birmingham)

The Aesthetic Appreciation of Perfumes as Multi-Functional Objects

I focus on an increasingly widespread practice of perfume appreciation that I term perfume sampling, in which perfume enthusiasts sample different scents for the sake of getting to know new interesting olfactory creations (see Brozzo 2020; Freeland 2022). I argue that, in the practice of perfume sampling, perfumes are aesthetically appreciated as multi-functional objects: perfumes are aesthetically appreciated as interesting olfactory creations, either in addition to, or independently of, their adornment function (see Davies 2020). In support of this idea, I first emphasise the way in which perfume enthusiasts will travel to some places (such as niche perfume stores, or institutions such as the Osmothéque) just to get acquainted with interesting scents, and not necessarily for acquiring them (scent acquisition is impossible at the Osmothéque, which is a scent archive). Secondly, I build on Larry Shiner's (2020) work, and I explore similarities between perfume sampling and the Japanese kōdō ceremony, where the scents of different pieces of incense wood are passed around among participants alongside short poems, and participants have the task of identifying these scents partly by means of paring them up with poems. I point out that, in the kōdō ceremony, the scents of pieces of incense wood passed around are aesthetically appreciated regardless of any practical application. Similarly, in perfume sampling, the aesthetic appreciation of perfumes is largely (though not wholly) independent of the aim of wearing them. This is the case although many of the perfumes being sampled could perfectly well be bought and worn (cf. Carlson and Parsons (2008) on cases of fashion design items that are difficult to wear). Thus, I uncover one aspect of the multi-functionality of perfumes beyond adornment (see also Martin-Seaver, ms., and Scarbrough, ms.).

References

Brozzo, C. (2020). Are some perfumes works of art?. The Journal of Aesthetics and Art Criticism, 78(1), 21–32.
Davies, S. (2020). Adornment: What Self-Decoration Tells Us about Who We Are. London: Bloomsbury.
Freeland, C. (2022). Art Scents: Exploring the Aesthetics of Smell and the Olfactory Arts. The Journal of Aesthetics and Art Criticism, 80(2), 248–251.
Martin-Seaver, M. (ms.). Perfume and Multi-function Design Objects.
Parsons, G., & Carlson, A. (2008). Functional Beauty. OUP Oxford.
Scarbrough (ms.). Our Perfumed Lives: Creating Scent Moods and Memories.
Shiner, L. (2020). Art scents: exploring the aesthetics of smell and the olfactory arts. Oxford University Press.

Michael Aaron Lindquist (Northland College)

Smelling the Good Life

This paper explores the various ways olfaction may fit into living "the good life," attending to the standard classification schema of such theories as constituting either (1) Hedonism, (2) a Desire-Satisfaction Theory, or (3) an Objective List Theory. While the olfactory sense is itself an aspect of well-being in a minimal way, the importance of which was recently highlighted by cases of COVID-19-related anosmia, olfaction through directed sniffing and attention may expand the role of olfaction in living a good life, and it is on the connection between attentive olfaction and the good life that the bulk of the paper focuses. Smell's role in hedonism differs with respect to the details of the particular variety, with this paper exploring applications of olfaction within both Benthamite and Millian hedonism. The subjectivist stance of Desire-Satisfaction Theories limits olfaction's role on the basis of the desires of a particular agents; discussions of what desires people should have under a desire-satisfaction theory may depend upon notions of coherency of a set of desires and the manner in which they may or may not mutually reinforce one another, but the subjectivism stance precludes a more substantive investigation. Objective List Theories, besides almost always including

happiness as a key component of the good life, often include other elements that smell neatly fits into, especially aesthetic appreciation. While much of the focus of aesthetic appreciation may be on art or nature, smell offers a further means of engaging with the aesthetics of the everyday. Further interrogating the role of smell in the aesthetics of the everyday uncovers the role of slowness in attaining the good life.

Giulia Martina (University of Tübingen)

Smell and misperception: The case of parosmia

How should we understand olfactory misperception? I explore this question by looking at a real-life example of altered olfactory experiences – parosmic experiences. Parosmic experiences can be very different in both sensory and hedonic character from the experiences of normal perceivers: wine, for instance, may smell revolting and like sewage. However, subjects' reports, phenomenology, and empirical research on the mechanisms of parosmia all support the idea that parosmic experiences involve some perceptual connection to the olfactory envrionment – they are not just smell hallucinations. I first question whether Batty's original theory of olfactory misperception can account for this idea. I will then explore whether the notion of property illusion introduced by Macpherson and Batty might help. Property illusions are defined as experiences that a) closely match, although not perfectly, the scene around the subject, and b) are appropriately caused by that scene. On the basis of recent empirical evidence, I argue that parosmic experiences might fail to satisfy this definition. If so, even this notion of misperception may not allow us to fully account for the nature of parosmic experiences. I propose that there are in fact two ways in which parosmic experiences, while altered, involve a perceptual connection to the environment: not only are subjects sensitive, albeit partially, to the olfactory qualities around them; they are also sensitive to the presence of odour sources (e.g. wine), which smell distorted and 'wrong'. An account that can appeal to both the role of olfactory qualities and the role of sources is thus better positioned to respect the distinctive features of parosmic experience.

Błażej Skrzypulec (Jagiellonian University)

Time, Actions, and Spatial Content of Olfaction

According to a common view, olfactory experiences lack well-developed spatial content. Nevertheless, there is also an important opposition to such a restricted perspective on olfactory spatiality which claims that a view ascribing only rudimentary spatial content to olfaction arises from a narrow focus on short and passive olfactory experiences. The goal of the paper is to investigate which, if any, version of such 'diachronic spatiality thesis' is plausible. The conducted investigations reveal three interesting results. First, even the truth of the most modest versions of the diachronic spatiality thesis relies on certain nontrivial assumptions regarding the temporal structure of consciousness. Second, while it is common to postulate that spatial content occurs in olfaction due to active sensing, in fact the actions themselves are of little significance for the olfactory spatial perception. In order to defend a postulate that active sensing significantly expands the olfactory spatial content, one has to also adopt a 'thick' view on olfactory content. According to such a view, the content of olfactory experiences is not obtained solely due to the functioning of mechanisms of orthonasal olfaction, but also in virtue of the functioning of other sensory systems. Third, while it is quite plausible that olfaction can represent a significant variety of egocentric spatial relations, it is less likely that the same is true about allocentric spatial relations between odors and odors' spatial properties such as size and shape. In consequence, there are serious doubts whether olfactory experiences can represent odors as positioned within smellscapes.

Benjamin D. Young (University of Nevada at Reno)

Can I smell YOU?

We can olfactorily perceive our own and others' odorprint, health, kinship, emotional status, psychological attributes, and social relation. Moreover, these perceptual states modulate our behavior even in the absence of conscious awareness (Young, forthcoming). Despite the robust nature of these capacities the talk explores the limits of human chemosignaling. In particular, empirical studies will be surveyed to examine if there is evidence that we can parse our own personal odor from those of others such that we can achieve a second-person olfactory perspective? The talk concludes by exploring if our limitations in smelling others' second personally is particular to social chemosignaling or attributable to general properties of olfactory processing such as its spatiality, vagaries of the sensory properties of olfactory objects, synthetic representational format, or lack of attentional modulation.