***The 14th London Reasoning Workshop: July 24-25, 2023***

***Room QA080, Queen Anne Court, Old Royal Naval College***

**Monday the 27th to Tuesday the 28th of July 2009**

**The 14th London Reasoning Workshop**

**Monday, July 24, 2023**

|  |  |
| --- | --- |
| **9:45** | **Welcome** |
| **10:00-10:30** | **Leon Schöppl**How complexly should we model the reasoning of Bayesian agents? |
| **10:30-11:00** | **Diana Osmolska**Architects’ use of intuition in site analysis: exploring the process of information gathering in architects’ solution development for the workshop  |
| **11:00-11:30** | **Coffee** |
| **11:30-12:00** | **Jonathan Evans**How chess engines changed human chess: A case study on the impact on artificial intelligence |
| **12:00-12:30** | **Igor Douven, Shira Elqayam & Nabil Hasshim *(Remote)***Inferentialism, Metacognition, and the Limits of Centering |
| **12:30-1:00** | **Samuel DeBray & Katrin Schulz**Conditional Reasoning with Causation in Large Language Models |
| **1:00-2:30** | **Lunch** |
| **2:30-3:00** | **Simone Sebben & David Over**Conjunctions of conditionals and their significance  |
| **3:00-3:30** | **Maxime Bourlier, Daniel Lassiter, Baptiste Jacquet & Jean Baratgin**How generics obscure the logic of conditionals  |
| **3:30-4:00** | **Laura Kelly & Sunny Khemleni**Iconicity Bias and duration |
| **4:00-4:30** | **Coffee** |
| **4:30-5:00** | **Corina Strößner & Ulrike Hahn *(Remote)***Learning from Conditional Probabilities  |
| **5:00-5:30** | **Peter Collins, Ulrike Hahn, & Karolina Krzyzanowska**Conditional reasoning and the framing of health communication |
| **5:30** | **Wine Reception** |

**Monday the 27th to Tuesday the 28th of July 2009**

**The 14th London Reasoning Workshop**

**Tuesday, July 25, 2023**

|  |  |
| --- | --- |
| **9:00-9:30** | **Omid Ghasemi, Simon Handley, & Stephanie Howarth *(Remote)***Reversing the Intuitive Logic Effects: The Role of a Matching Heuristic in Argument Evaluation |
| **9:30-10:00** | **Vinod Goel**Blended Response Hypothesis  |
| **10:00-10:30** | **Nick Chater & Mike Oaksford**Spontaneous Rationality  |
| **10:30-11:00** | **Coffee** |
| **11:00-11:30** | **Calvin Deans-Brown and Henrik Singmann**The belief bias in everyday political reasoning: evidence from a US sample  |
| **11:30-12:00** | **Cillian McHugh, Marek McGann, Eric Igou, & Elaine Kinsella (*Remote)***Distancing and Moral Dumbfounding; Facilitating the Identification of Reasons |
| **12:00-12:30** | **Niels Skovgaard-Olsen & Christoph Klauer**Norm Conflicts and Morality: The Conflict Model of Moral Decision-Making |
| **12:30-2:00** | **Lunch** |
| **2:00-2:30** | **Marvin Law, Sabina Kleitman, Lazar Stankov, & Valerie Thompson** Giving up: How, why and so what?  |
| **2:30-3:00** | **Rakefet Ackerman & Yoed Kenett *(Remote)***Biasing Factors for Originality Judgments  |
| **3:00-3:30** | **Valerie Thompson**What makes us think? |
| **3:30-4:00** | **Coffee** |
| **4:00-4:30** | **Katharina Zimmermann & Nicole Cruz (*Remote)***Jumping to Racial Prejudice  |
| **4:30-5:00** | **Nick Byrd (*Remote)***Reflection-Philosophy Order Effects and Correlations: Aggregating and Comparing Results of mTurk, CloudResearch, Prolific, and undergraduate samples |

**Abstracts**

In order of presentation

**Laura Macci & Maria Bagassi (withdrawn)**

The interpretive function in insight problem solving

Language and thought share a unitary cognitive activity, addressed by an interpretative function. This interpretative effort reveals the assonance between the attribution of meaning to an utterance and the discovery of a solution via restructuring in insight problem solving. We suggest a view of complex integrated analytical thinking, which assumes that thinking processes information in different ways, depending on the characteristics of the tasks the subject has to solve, so that reasoning results in a stepwise, rule-based process or in a widespread activity of search where implicit parallel processes are also involved (unconscious analytic thought). We investigated the interrelationship between language and thought in insight problem solving, in both its positive (Experiments 1 and 3) and its negative effects (Experiment 2). Our results are discussed in the light of the debate on dual processing theories.

**Laura Caravona, Francesco Poli & Laura Macchi (withdrawn)**

Dissociable effects of verbalization on solving insight and non-insight problems

Numerous studies have investigated verbalization effects with the aim of clarifying the processes responsible for solving insight and non-insight problems. However, there is no concordance in the literature among the various theories. While studies supporting the *special process* view have reported different effects of verbalization in insight and non-insight problems, highlighting an overshadowing effect in the former, studies accounting for the *business-as-usual* approach reported a lack of verbalization effect in insight problems and claimed that the solution processes of the two types of problems are more similar than different. The study investigates, with the same procedure, the effect of verbalization on insight and non-insight problems, providing evidence supporting the role of unconscious processes for the resolution of the former, in contrast with the stepwise, conscious procedure for the resolution of the latter. These results will be discussed in the light of a third approach, Unconscious Analytic Thought (Macchi & Bagassi, 2012, 2015; Bagassi & Macchi, 2016), which claims that the insight problem-solving process is mainly unconscious and analytical.

|  |
| --- |
| **Leon Schöppl**How complexly should we model the reasoning of Bayesian agents? |

Bayesian source-reliability models provide a way to encode the likelihood ratios required for conditionalization on testimony from an advisor about a hypothesis. They are simple Bayesian networks specifying the probabilistic dependencies between (1) the truth of the hypothesis,
(2) the reliability of the source, and (3) their giving testimony as to that hypothesis being true. These models can be employed in network simulations as stand-ins for the psychology of boundedly rational
reasoners, which allows the computational study of large-scale phenomena relevant to social epistemology, like polarization and the formation of echo chambers or filter bubbles.

However, the literature proposes a variety of these source-reliability models, which differ greatly in their respective complexities: (1) they range from merely representing very limited cases of source-reliability to capturing the entire spectrum, (2) not all of them feature nodes to explicitly encode an advisor's honesty or alignment, and (3) many of them cannot provide reliability assessments that are

sensitive to the amount of information available about the source.

Using a non-Bayesian model of the testimonial situation, I compare and contrast some of these models to spark a discussion about the tradeoffs between limiting computational complexity and honoring an adequate notion of bounded rationality. How detailed do they have to be to allow the drawing of conclusions about real-world reasoning?

**Diana Osmolska**

Architects’ use of intuition in site analysis; exploring the process of information gathering in architects’ solution development for the workshop

This study explores site analysis, which involves designing a building in outline to test whether a particular site can accommodate a given proposal. During this process, information is gathered to develop and test solution-conjectures. Designers' examination of solution-conjectures is underexplored. In addressing this issue, the study draws on dual-processing theory, where cognitive processes are divided into Type 1 and 2; the former being intuitive and efficient, and the latter effortful and slow. 21 interviews were conducted. The findings indicate that architects can avoid complex analysis by reframing difficult questions to reach solution-information-satisfaction. The scale of confidence associated with narratives accompanying architects’ solution development affected the detection of missing information, prompting solution-satisfaction on proposals with inadequate supporting information, leading to information neglect, major delays, and other problems.

**Jonathan St. B. T. Evans**

How chess engines changed human chess: A case study on the impact of artificial intelligence

Computers have changed human chess in several ways, but the most dramatic impact is from the advent of AI chess programs, or engines, which can now play the game much better than the top human players. Interestingly, this has not reduced interest in watching the top human players whose tournament play is now regularly streamed live on the internet with expert commentary for chess fans. By contrast, there is little interest in bot on bot encounters, even though the standard of play is much higher. Engines have actually strengthened human play, as players can now analyse their own games and those of others more accurately, and also evaluate potential moves and positions that have never been played before. Professional players now routinely use engines to prepare surprise opening variations, for example. Specialists also study the games played by the top-rated programs, especially those using neural nets, to improve human understanding of the game. However, chess engines do not convey the difficulty of chess positions for human players, for reasons I will explain. The main negative impact is on the potential for cheating, when even a smartphone with the right software can beat a world champion, and ever more elaborate precautions are being taken to prevent or detect this. Finally, I discuss how I think chess engines will change in the near future, and why this will make it harder to detect whether players are cheating.

**Igor Douven, Shira Elqayam, & Nabil Hasshim**

Inferentialism, Metacognition, and the Limits of Centering

According to Inferentialism, a conditional is true precisely if there is a compelling argument for its consequent starting from its antecedent conjoined with contextually given background knowledge. As a direct consequence of this view, the truth of a conditional's component parts does not guarantee the truth of the conditional, given that it is compatible with there not being an argument from the conditional's antecedent to its consequent. Thus, Inferentialism violates a rule called 'Centering', which licenses the inference of "If A, B" from the joint truth of A and B.  This distinguishes Inferentialism from all standard semantics of conditionals, each of which validates Centering. Here, we report the results from a study that registered (i) participants' truth evaluations of a set of conditionals; (ii) their truth evaluations of the component parts of those conditionals; (iii) their judgments of the strength of the inferential connection between antecedent and consequent for each conditional; and (iv) their metacognitive confidence in the inference strength judgments. The data from (i)-(iii) allowed us to test Centering and to investigate whether possible violations of the rule can be explained in inferentialist terms. In combination with the data from (iv), this also allowed us to test psychological predictions derived from Hypothetical Inferential Theory, the psychological face of Inferentialism. Our main finding is that people tend to obey Centering only for conditionals that, from their perspective, satisfy the truth condition of Inferentialism.

**Samuel Debray & Katrin Schulz**

Conditional Reasoning with Causation in Large Language Models

Causal conditionals have been widely studied in the psychology of reasoning literature. However, very little has been done to assess how well state-of-the-art language models (such as GPT-4, LLAMA, OPT, etc.) handle inferences with causal conditionals. We crowd-sourced a large dataset of conditionals involving different kinds of causal relations between the antecedent A and the consequent B (namely A causes B, B causes A, A and B are independent, and A and B are jabberwocky sentences for control). First, we used this dataset to verify, during an online behavioural experiment, that the causal relation between A and B has a strong effect on human reaction time when judgement conditional inferences as valid or invalid. Ongoing work focuses on how well Meta AI’s OPT-66B can predict the validity of such conditional inferences, and whether its performance also depends on the causal relation between A and B. We are also trying to find out whether there exists some correlation between the model’s surprisal and human reaction times when judging these inferences.

**Simone Sebben & David Over**

Conjunctions of conditionals and their significance

There are long-standing disagreements in logic and philosophy about how to analyse conjunctions of conditionals in natural language (Cantwell, in press; Edgington, 1991; Lance, 1991; McGee, 1989; Sanfilippo et al., 2020). The psychology of reasoning has totally ignored this topic, but we will make a case for conducting empirical research on it and present for the first time data relevant to it. Of particular interest are conjunctions of the form *if A then C & if not-A then C*. These conjunctions can be used to "sum up" the logically valid inference of Dilemma: inferring *C* from a derivation of *C* from *A* and a derivation of *C* from *not-A*. A normative analysis of the form *if A then C & if not-A then C* must have the result that *C* validly follows from it. Dilemma is equivalent as an inference to double negation, classical reductio ad absurdum, and excluded middle. Another function of this form is to convey that *C* is independent of *A*, making both *if A then C* and *if not-A then C* "independence conditionals" (Cruz & Over, 2023). We will present an example in which, according to Edgington and Lance, the probability of *if A then C & if not-A then C* is 0.5, according to McGee and Sanfilippo et al., this probability is 0.25, and according to strong truth condition inferentialism, it is 0. We will discuss how well these competing predictions match our empirical data. We will also argue that a prediction about this probability cannot be derived from recent revisions of mental model theory or from weak truth condition inferentialism, but we will be, of course, open to correction.

**Maxime Bourlier, Daniel Lassiter, Baptiste Jacquet & Jean Baratgin**

How generics obscure the logic of conditionals

Indicative conditionals containing modals and adverbs of quantification generate apparent counter-examples to core logical principles such as modus ponens (MP) and modus tollens (MT). For instance, from “If Jim has skipped breakfast, Bill is usually annoyed” and “Jim has skipped breakfast”, we cannot conclude “Bill is usually annoyed”. (Imagine, in particular, that Jim rarely skips breakfast, and nothing else annoys Bill.) Does this mean that a vast psychological literature showing the validity of modus ponens is mistaken? No: as linguists have shown, both the antecedent and consequent of the conditional receive a distinct interpretation from their unembedded instances. We can construct similar apparent counter-examples to MT: from the same conditional and “Bill is not usually annoyed”, we cannot conclude that Jim ate breakfast.

Here we can pinpoint the issue in the interaction of “usually” with the “if”-clause. The literature on default reasoning gives us similar failures of modus ponens in examples with a generic interpretation with no overt operators to blame, such as the Tweety paradox. These have motivated popular MP-denying treatments of conditionals using  “defeasible implication”, with a close relative in Kratzer’s influential theory. We argue that it is a mistake to base conclusions about the logic of conditionals on such apparent failures without carefully controlling for the presence or absence of generic and habitual language. These sentences contain implicit operators analogous to “usually” that interact with conditional antecedents to create illusory counter-examples to basic patterns of conditional reasoning. To control for such interactions, researchers should focus on reasoning with sentences that enforce a single-event, episodic or stative interpretation with no implicit quantification over cases.

As a case study, we analyze examples proposed by Douven (2016) as problem cases for the conditional semantics of de Finetti (1936), involving conditionals that embed other conditionals in their antecedents and/or consequents. While the argument is initially compelling, the purported counter-examples are uniformly generic or habitual in form. We present the results of an experimental study that explicitly manipulates the presence of generic vs. episodic language in nested conditionals. Participants’ judgments adhere closely to de Finetti’s predictions for episodic conditionals, and pattern in generic conditionals as we would expect when combining de Finetti’s theory with linguistic theories of genericity. The results illustrate the importance of avoiding generics in experimental studies of conditionals, and more generally of careful attention to linguistic detail when constructing and analyzing reasoning experiments.

**Laura Kelly & Sunny Khemlani**

Iconicity bias and duration

Abstract: Descriptions of durational relations can be ambiguous, e.g., the description ‘one meeting happened during another’ could mean that one meeting started before the other ended, or it could mean that the meetings started and ended simultaneously. A recent theory posits that people mentally simulate descriptions of durational events by representing their starts and ends along a spatial axis, i.e., an iconic representation of time. To draw conclusions from this iconic mental model, reasoners consciously scan it in the direction of earlier to later time points. The account predicts an iconicity bias: people should prefer descriptions that are congruent with an iconic scanning procedure – descriptions that mention the starts of events before the ends of events – over logically equivalent but incongruent descriptions. Six experiments corroborated the prediction; they show that iconicity biases in temporal reasoning manifest in cases when reasoners consciously evaluate the durations of events.

**Corina Strößner & Ulrike Hahn**

Learning from conditional probabilities

Conditional probabilities are at the core of Bayesianism. They play a central role in the learning method of conditionalisation and the application of Bayes' theorem. Although conditional probabilities are central to Bayesian learning models, little is still known about how well human agents can learn from conditional probabilities as evidence. This paper addresses this question in a theoretical and experimental way.

In the theoretical part, we distinguish between cases of vacuous learning, where the probabilistic information learned is not new and belief revision is not warranted, and cases with genuinely new information. We also discuss the role of conservative constraints, which require an agent to change its belief only as much as is necessary to accommodate the new evidence.

In the experimental part, we examine how well human intuition fits to probabilistic norms, adding new insights to the long-standing question of the extent to which the human mind is adapted to probabilistic norms. We report an exploratory study in which we found a moderate fit between human subjects and our normative model. In a larger second study with more material and a stronger focus on the inference task, we found high agreement between human judgement and the normative model. This gives reason to believe that humans (at least from the educated, Western population we tested) are well adapted to correctly solve probabilistic reasoning tasks when the evidence is in the form of conditional probabilities.

**Peter Collins, Ulrike Hahn, & Karolina Krzyzanowska**

Conditional reasoning and the framing of health communication

Should a health campaign emphasise the potential gains from compliance (e.g., ‘If you give up smoking, you’ll reduce your risk of lung cancer’) or the potential losses from non-compliance (e.g., ‘If you don’t give up smoking, you won’t reduce your risk of lung cancer’)? A large literature on so-called goal framing, or message framing, holds that such messages are logically equivalent but should be persuasive in different contexts (see, e.g., Gallagher & Updegraff, 2012). For example, according to the risk-matching hypothesis (Rothman & Salovey, 1997), gain frames should be used with behaviours perceived to be low in short-term risk, and loss frames should be used with behaviours perceived to be high in short-term risk. As the risk-matching hypothesis has struggled to find support, calls have arisen for new theoretical frameworks (Nan et al., 2018; van’t Riet et al., 2016). We argue for an alternative approach grounded in research on conditionals, the linguistic vehicle of goal frames. We dispute the assumption that frames necessarily convey logically equivalent information and argue that frames must be interpreted in context to understand their persuasive effects. We make a first step towards this alternative approach and outline an empirical study which investigates the effects of disablers and alternative causes on participants’ response to frames.

**Omid Ghasemi, Simon Handley, & Stephanie Howarth**

Reversing the Intuitive Logic Effects: The Role of a Matching Heuristic in Argument Evaluation

The capacity to evaluate logical arguments intuitively is a fundamental assumption of recent dual-process theories. One observation supporting this effect is the standard conflict effect on incongruent arguments under belief instruction. Conflict arguments are evaluated less accurately than non-conflict arguments, arguably because logic is intuitive and automatic enough to interfere with belief judgments. However, recent studies have challenged this interpretation by finding the same conflict effects when a matching heuristic cues the same response as logic, even on arguments with no logically valid structures. In this study, we test the matching heuristic hypothesis across 4 experiments (total *N* = 409) by manipulating the arguments propositions so that *matching* cues a response that is either (1) aligned or (2) misaligned with logic, or (3) cues no response at all. Consistent with the predictions of the matching heuristic, standard, reversed, and no conflict effects were found in those conditions, respectively. These results indicate that intuitively correct inferences which are assumed as evidence of logical intuitions are actually driven by a matching heuristic that cues responses aligned with logic. Alleged intuitive logic effects are reversed when the matching heuristic cues an opposing logical response or disappears when there are no matching cues. Therefore, it appears as though the operation of a matching heuristic, rather than an intuitive access to logic, drives logical intuitions.

**Vinod Goel**

Blended Response Hypotheses

A great deal of effort has been expended over the past decades arguing over a single system versus a dual system reasoning model.  Within the dual system it is always one or the other system that responds.  This paradigm may work for some artificially contrived laboratory problems (e.g. the Linda problem), but  when one looks more broadly in the world, it becomes apparent that human behavioral responses are blended responses of multiple systems.  I will review some data from the literature (e.g. economic decision-making tasks, discriminative parental solicitude effect) and some data we are currently collecting in our lab to make the case for the blended response hypothesis, and offer a model of tethered rationality that can accommodate it.

**Nick Chater & Mike Oaksford**

Spontaneous Rationality

This talk presents the case for spontaneous rationality: that distributed, local, and largely public, processes of debate, reflection, action and observation can lead to (but do not *justify*) increasingly rational thoughts, theories, markets and behaviours. Spontaneous rationality emerges from the incoherence and paradox of everyday thought---consistency is rarely achieved, and only as the end-point of reasoning, not a precondition for reasoning. Spontaneous rationality typically arises from adversarial interactions: each agent seeks to exploit local incoherence in others, and to find, and resolve, local incoherence in themselves. This viewpoint on rationality resolves a variety of puzzles within the cognitive sciences; and has far-reaching implications for folk psychological explanation, epistemology, and formal models of reasoning in the cognitive sciences.

**Dean-Brown**

The belief bias in everyday political reasoning: evidence from a US sample

During the COVID-19 pandemic we saw a rapid proliferation of information, including misinformation and fake news, which has been termed an ‘infodemic’. Much of this information was spread by online discourse, where we saw that many people were persuaded by even poorly reasoned arguments. It is of interest as scientists to understand why some people are persuaded by arguments of poor quality, while others are not.

Evans and colleagues began to answer this question in 1983, with an experiment investigating the role of belief on argument evaluation. In this experiment that has been replicated numerous times, participants’ task is to judge whether the conclusion of presented syllogisms (a form of logical argument) follows from the syllogisms’ premises with logical necessity. Syllogisms are manipulated with regard to their validity (valid vs invalid) and the believability of their conclusions (believable vs unbelievable). Whereas participants can distinguish valid and invalid syllogisms (albeit imperfectly), participants are also more likely to rate syllogisms as logically valid if the syllogism has a believable conclusion compared to an unbelievable conclusion. Additionally, the effect of validity is stronger for syllogisms with unbelievable conclusions compared to syllogisms with believable conclusions. This pattern of findings is known as the belief bias.

Syllogisms are formal reasoning schemas that are unlike arguments we typically see in everyday ‘real-world’ argumentation. We present two experiments that investigate the belief bias effect in the context of everyday arguments, such as those from newspapers and social media. The arguments in our study differ in terms of their informal argument quality; ‘good’ arguments provide an explanation for their conclusion, whilst ‘bad’ arguments do not provide an explanation and contain a common fallacy (e.g., circular reasoning, arguments from authority).

In our experiments, participants rate their beliefs about a series of political claims (e.g., ‘abortion should be legal’) and rate the strength of ‘good’ or ‘bad’ arguments about these claims. Our results find the belief bias for everyday arguments; whilst participants consistently rate good arguments as stronger than bad arguments, they are also biased in rating arguments in line with their beliefs as stronger than arguments that are not. In the second experiment, we replicate and strengthen this finding. Participants still use their beliefs to judge argument quality even when told explicitly not to, regardless of whether the arguments are presented before or after measuring participants’ beliefs. With regards to combating misinformation, our findings suggest that combating poor quality information consistent with someone’s beliefs with good quality information inconsistent with their beliefs is ineffective, as belief-inconsistent information is generally seen as worse than beliefconsistent information.

**Cillian McHugh, Marek McGann, Eric Igou, & Elaine Kinsella**

Distancing and Moral Dumbfounding: Facilitating the Identification of Reasons

Moral dumbfounding occurs when people maintain a moral judgment in the absence of supporting reasons. Drawing on dual-process approaches to moral judgment, one possible explanation for moral dumbfounding proposes that it occurs as a result of a conflict between intuitive and deliberative processes. Consistent with this explanation, previous research has shown that under manipulations designed to lead to more intuitive thinking rather than deliberative thinking (such as increased cognitive load), people are less likely to provide reasons for their judgments, and more likely to provide dumbfounded responses in a moral dumbfounding task. Building on this work the present research examines if dumbfounded responding can be reduced through experimental manipulations designed to facilitate deliberative thinking (over intuitive thinking). Drawing on construal-level theory, and the finding that distancing manipulations (e.g., psychological/temporal distancing etc.) facilitate deliberative thinking, we predict that including a distancing manipulation in a moral dumbfounding task will increase reason-giving, and reduce dumbfounded responding. We present eight pilot studies testing the feasibility of different distancing manipulations, with inconclusive results. We propose a high-powered pre-registered study to test this prediction.

**Niels Skovgaard-Olsen & Christoph Klauer**

Norm Conflicts and Morality: The Conflict Model of Moral Decision-Making

The goal of this paper is to study individual variation in participants’ adherence to conflicting moral views. To do this, we introduce a new conflict model of moral judgment which builds on the widely used CNI model of moral judgments (Gawronski et al., 2017) but improves it in several aspects. First, we follow Skovgaard-Olsen and Klauer (2023) in extending the model to investigate invariance violations of the models’ parameters. Second, we extend the CNI model to permit cases where participants are conflicted between utilitarian and deontological response tendencies. Third, we extend the model to investigate how the difference between permitting instrumental as a foreseeable side-effect and as an intended means affects the model parameters. Fourth, we employ an additional experimental paradigm to estimate latent classes in participants’ moral views and the new Conflict Model to estimate group-specific means in the model parameters. Finally, we measure a range of egoistic and altruistic covariates used in Kahane et al. (2015) and Conway et al. (2018) to investigate whether participants’ acceptance of instrumental harm is associated with a genuine concern for the greater good or whether it is rather driven by asocial character traits (Bartels & Pizarro, 2011).

**Marvin Law, Sabina Kleitman, Lazar Stankov, & Valerie Thompson**

Giving up: How, why and so what?

Under the Meta-reasoning (MR) model, the decision to give up reflects a metacognitive strategy that can mitigate errors and resource costs when solving problems. However, there is a need for an improved understanding of how individuals give up, whether such strategies are used systematically and whether they are adaptive or maladaptive for performance. In two studies (N1=176 & N2=370), we examined this behaviour from individual differences and cognitive perspectives. Time spent until giving up, frequency of giving up, cognitive and metacognitive indices embedded within cognitive tests, as well as decision-making styles and performance metrics were recorded. In both studies, participants varied systematically in their giving-up behaviour: frequencies converged onto a Giving Up factor using Confirmatory Factor Analysis. Participants also clustered into profiles based on time spent until giving up and frequency. Using Latent Profile Analyses in Study 2, three unique giving-up profiles emerged—1) fast and frequent, 2) fast and rare, and 3) slow and rare. The two “fast” profiles were characterised as maladaptive, having lower points and bets than the adaptive “slow” profile. Mixed-effects linear regression indicated that people gave up more frequently when feelings of confidence were low and faster when a problem was perceived as unsolvable, indicating that judgements of solvability (iJOS) and confidence judgements are critical to giving-up decisions during problem-solving.

The results suggest that individuals have trait-like tendencies to give up and are systematically predisposed towards adaptive and maladaptive giving-up strategies. Maladaptive profiles prefer fast giving-up decisions, and adaptive profiles prefer to give up slowly. Overall, our results validate the processes proposed by the Meta-Reasoning framework and provide a strong foundation for further investigation into monitoring processes and the role of giving up in decision-making. Practical implications for learning and problem-solving and future directions are discussed.

**Rakefet Ackerman & Yoed Kennet**

Biasing Factors for Originality Judgments

Is my idea creative? This question pertains to investing in companies and selecting research agendas. Building on previous research, our focus is on the originality of ideas and how they relate to the self-assessments of the individuals generating those ideas. We operationalize originality score as the frequency (%) of each idea within a sample of participants and originality judgment as the self-assessment of this frequency. Preliminary evidence suggests that originality scores and originality judgments are the products of distinct processes. Consequently, originality judgments are susceptible to biases. However, the specific heuristic cues that contribute to these biases are not well-known. To address this gap, we employed meta-reasoning methods to identify the heuristic cues underlying originality judgments. Additionally, we utilized computational linguistics to examine semantic distance in a generalized association net as a potential source of bias. Our findings reveal that both traditional cues and semantic distance underly originality judgments. This research provides insights into uncovering metacognitive biases in ill-defined tasks and suggests that semantic distance serves as an unrecognized metacognitive cue.

**Valerie Thompson**

What Make Us Think?

Valerie A Thompson, University of Saskatchewan; Valerie.thompson@usask.ca

Metacognitive processes monitor and control first-order cognitive processes. Their goal is to assess how well the processes are doing (e.g., I’m not certain that I have remembered the person’s name correctly) and to control subsequent processing (e.g., don’t address them by name and hope it comes to you). In the domain of reasoning, one of the most important control processes is to determine when to think more deeply about a topic or to terminate processing. In this talk, I propose to integrate two metacognitive models of how people control their reasoning: Stanovich’s Rational Mind Model and Thompson’s Feeling of Rightness approach. In this way, I propose that we may understand the answer to two broad questions: 1) why, all other things being equal, do some people think more than others, and 2) why all other things being equal, does a person think more about some things than others?

**Katharina Zimmermann & Nicole Cruz**

Jumping to racial prejudice

Jumping to conclusions (JTC) refers to the tendency to reach a conclusion or decision without sufficient evidence to justify it, similar to an unwarranted generalization. It has been investigated mostly in people with schizophrenia, but has also been found to correlate with conservative political beliefs and with conspiracist thinking. Taking as a starting point recent reports of racist violence in the United States, we report two online experiments investigating whether jumping to conclusions tends to be higher in people with stronger racial prejudice as well as with stronger scores in the related constructs of need for closure, social dominance orientation, nationalism, and rejection of science. In Exp 1, US participants completed the classic beads task with neutral content (e.g. red or blue beads drawn from a jar). In Exp 2, the same participants worked through an adaptation of the beads task to contents relevant to racist and political beliefs (e.g. local or foreign crime suspects drawn from a police department). Overall, we found that racist beliefs, and among the constructs related to racist beliefs in particular need for closure, were associated with higher JTC independently of the task content. Further, JTC was higher for content congruent with racist beliefs in people who scored higher on racism. Follow-up work will assess whether the trial by trial subjective probability judgments associated with the task are calibrated with Bayesian principles or suggest over- or underconfidence in participants’ decisions. The present findings are unfortunately limited to the US context, but could be extended to other geographic and cultural contexts in which racial prejudice occurs.

**Nick Byrd**

Reflection-Philosophy Order Effects and Correlations: Aggregating and Comparing Results of mTurk, CloudResearch, Prolific, and undergraduate samples

Failed replications sometimes elicit speculation about differences in participants or data quality. To test such explanations of mixed correlational and causal results regarding reflective and philosophical thinking, one pre-registered experiment recruited 460 participants from mTurk, CloudResearch, Prolific, and a university. A classic reflection test prime randomized participants to start with either reflection tests or philosophical thought experiments. First- and thirdparty data quality measures found up to 18 times more low-quality submissions in the mTurk sample. Among acceptable submissions (n = 385), some correlations between reflective and philosophical decisions replicated (e.g., denying accidentally true belief is knowledge), but no reflection test priming effects replicated. Unexpectedly, a philosophical reflection effect emerged: making philosophical decisions before the reflection tests slightly improved reflection test performance (only in the acceptable submissions). This challenges the proposed direction of causation between reflective and philosophical thinking, but affirms concerns about replicability depending on participant source and data quality