SNE 2024 Abstract Proceedings

Table of Contents

Invited Speakers

Kavli Plenary Lecture

Algorithms that Understand and Understood by People; and Misunderstood and Misunderstood by People

Sendhil Mullainathan, *MIT*

I will argue there is a deep synergy between behavioral science and algorithmic science. We can do better behavioral science by using algorithms; and we can build better algorithms by relying on behavioral science. Exploiting that synergy though will require some changes: (i) building of new algorithmic tools for behavioral scientists; and (ii) building models of human behavior that can be incorporated into computational pipelines. I will illustrate these admittedly abstract points through several concrete projects.

Workshop I: Neuroscience

Changing the choices that we take during decision making; interactions between frontal cortex and neuromodulatory systems

Matthew Rushworth, *University of Oxford*

Every day we make decisions about what we will do and, even more fundamentally, whether it is worth doing anything at all. Often, when we have the opportunity to decide what to do or whether to act at all, we take the same choice as we did the last time. Sometimes, however, we strike out on a new course of action or change our mind about our policy for what is or is not worth doing. I consider how patterns of activity that play across the anterior cingulate cortex (ACC) and the dorsal raphe nucleus (DRN: the source of the serotonergic system) in people and monkeys are related to persisting in making the same choice or switching to an alternative. In addition, I show that it is now possible to use focal transcranial ultrasound stimulation (TUS) to selectively manipulate activity in a minimally invasive manner even deep in the brain in areas such as the ACC and the DRN to alter which choices are made. Manipulating activity in cortical regions such as ACC and subcortical regions such as DRN produces quite different patterns of change in flexible decision making and decision policy setting.

Workshop II: Social and decision science

Reconsidering the Path for Neural and Physiological Methods in Consumer Psychology

Hilke Plassmann, *INSEAD*

Recent decades have witnessed a burst of neuroscience research investigating mental and physiological processes that are central to consumer behavior, including sensory perception, memory, and decision-making. Nonetheless, few publications that include neural and physiological measures, or develop conceptual frameworks around neuroscience principles, have been published in consumer psychology. In this presentation I will discuss the reasons for this development and what it implies for future research in this area. I conclude with recommendations for how a new path for consumer neuroscience can advance a multidisciplinary approach to both understanding the consumer and empowering their welfare. There will also be time to discuss some practical tips if you (or your students) would like to apply for a faculty position at a Business School.

Symposia

S.01: Ecological approaches to social learning and decision-making

Symposium Chair: Patricia Lockwood, University of Birmingham

Humans are ultrasocial - strongly influenced by what others think and do. Psychologists, neuroscientists, and economists have therefore sought to understand the neural and computational mechanisms that drive our complex social interactions. Whilst progress has been made, social behaviour does not occur in a vacuum. Instead, our social interactions are multidimensional and influenced by the environment, groups, and hierarchies in which they take place. This symposium will bring together leading neuroeconomists studying social learning and decision-making who have applied a diversity of ecological approaches to examine human social behaviour.

S.01.01: Features and dynamics of social inference and decision-making in naturalistic human interaction

Tessa Rusch [1], John O'Doherty [2], Ralph Adolphs [2], Nina Rouhani [1]
[1] Caltech, [2] California Institute of Technology

Details:
The ability to anticipate and react to others' behaviors is essential for adaptive social interaction. While many theories suggest that social behavior is influenced by inferences about hidden personal characteristics such as personality traits and emotional states, verifying these theories poses significant challenges due to the private nature of both the characteristics and

Back to top

the inference processes involved. In this study, we investigated whether and how inferences about a social partner's traits and states affect interactions with that individual. We analyzed the dynamics of social inference as pairs of participants engaged in naturalistic, unconstrained text-based conversations, followed by an iterative decision-making game where they competed or collaborated for monetary rewards (N=292). In two pilot studies (N=720), we developed a data-driven method to measure social inference. Our findings indicate that both chat and game interactions informed these inferences and significantly influenced their dynamics. Initially, there was a strong positivity bias in how participants evaluated their partners post-conversation. However, this bias diminished after the game, with participants adopting a more negative view of their partners. To determine whether the inferences drawn from conversations directly influenced decisions in the game, we applied computational cognitive models to the decision task data. We contrasted models that incorporated social inferences into the decision-making processes with those that relied solely on the history of the partner's choices. The best-fitting model integrated individual inequity aversion with adaptive learning about a partner's characteristics. Additionally, trait estimates generated through these models predicted the observed shifts from positive to negative trait evaluations following the game. These results not only support the theory that social inferences guide interactive behavior but also demonstrate the utility of computational models in deciphering complex social interactions.

S.01.02: Testing a neural autopilot theory using social media and power law estimation

Colin Camerer [1]

[1] *Caltech*

Details:

The neural autopilot model of habit formation describes transitions between low-cost habits and more effortful goal-directed choice. It is used as a structural (or generative) model to estimate behavioral parameters governing individual daily posting on the Chinese site Weibo around the 2020 Wuhan lockdown period by N=1272 users. The study's objective is to test a neurally-inspired theory in a large social ecology. The result of the structural estimation indicates that posting is generally goal-directed. Statistical persistence of choice (due to preference change, inertia or perseveration) adds a little predictive power to the autopilot structure. The model can be used to judge whether posting habits formed during the lockdown period, and persisted afterward, but does not show such "forced experimentation" learning.

Counterfactual simulations imply that artificially reducing the volatility in posting rewards would have significantly increased habitual posting. This implication is consistent with the autopilot premise that predictability of reward is part of habit formation. The results are important

because they show that neurally-inspired models can be applied to social ecology data, and invite application to many other human activities that become habitual.

The same neural autopilot model shows promise for explaining power laws in field data. The power law P(t)=ct-a

relates the frequency P(t) of an inter-activity interval t (e.g. 3 days between gym visits) to the exponentiated value of t. Power laws often fit behavioral data extremely closely (R2».95). A prominent mathematical foundation for power laws is variants of “priority queueing”, in which behavioral ‘arrivals’ are governed by a Poisson process which is sometimes overridden by urgent priorities. Neural autopilot simulations can approximate power laws that are evident in four new data sets on cellphone use, teacher educational software use, gym attendance, and rideshare work.

S.01.03: Dynamics of Representational Geometry in Social Hierarchies and Cognitive Control
Seongmin Park [1], Maryam Zolfaghar [2], Jacob Russin [3], Douglas Miller [4], Randall O'reilly [3], Erie Boorman [5]
[1] CNRS, [2] Department of Computer Science, UC Davis, [3] Center for Neuroscience, UC Davis, USA, [4] Center for Mind and Brain, UC Davis, USA, [5] University of California, Davis

Details:

The brain abstracts non-spatial relationships, such as social networks, into cognitive maps. Recent studies have demonstrated that the brain constructs a multidimensional representation of social hierarchies by integrating experiences from individual social interactions across various dimensions of these hierarchies. However, it remains unclear how the brain accesses information in the goal-relevant dimension from these maps when task goals continuously change. In this study, we explored how cognitive control facilitates the flexible selection of relevant information from multidimensional social cognitive maps stored in memory. We investigated the interplay between cognitive control and representational geometry using parallel analyses of human fMRI data (n=27) and recurrent neural network (RNN) models trained on identical tasks. Our findings reveal stable, map-like representations in the medial temporal lobe and orbitofrontal cortex, which encompass both relevant and irrelevant task dimensions. Additionally, we observed dynamic and orthogonal representations of only the relevant task dimensions, effectively compressing the irrelevant dimensions in the frontoparietal network. These representational patterns also displayed distinct temporal dynamics during RNN training, with map-like representations emerging first. Furthermore, we found that increased control demands, triggered by incongruences between relevant and irrelevant task dimensions, significantly altered the geometry of neural representations of

cognitive maps. These alterations in representation geometry correlate with individual differences in cognitive control capabilities and further suggest that distortions in the representation of multidimensional cognitive maps may underlie individual differences in decision-making biases. Our results highlight how different representational geometries maintain a balance between stability and behavioral flexibility and underscore a complex, bidirectional relationship between cognitive control and cognitive map geometry.

S.01.04: Ecological influences on human prosocial decision-making

Patricia Lockwood [1]

[1] *University of Birmingham*

Details:

Prosocial behaviours—decisions that help others—are critical for solving global challenges including climate change, infectious disease, global conflict, and ageing populations. So far, these behaviours have been studied with economic games or tasks where people decide between helping or not. However, theories from ecology suggest that there are important influences of the environment on when animals decide to act, yet how these ecological influences shape prosocial behaviours remains unknown.

I will discuss two projects that draw on fundamental principles from ecology - prey-selection and producer-scrounger dynamics - to understand human prosocial behaviour. In the first project, 510 participants across 3 studies completed a paradigm where they had to decide whether to interrupt their ongoing behaviour (watching a movie) to put in effort to reward another person, or on separate trials, to reward themselves. In poor environments, average reward values and probabilities were lower. In rich environments, average reward values and probabilities were higher. We demonstrate a robust ecological influence on decisions that benefit others. People were more willing to interrupt their behaviour to help others in poor compared to rich environments, and this influence was stronger compared to decisions that benefitted oneself. Computational modelling revealed that the opportunity costs of different environments were encoded distinctly for others and self. Value sensitivity when deciding for others in poor environments reached the same level as decisions to help oneself in rich environments.

In the second project, 200 participants in 35 groups interacted live and decided on each trial whether to produce to benefit the group or to free ride and collect rewards without contributing. Unlike a traditional public goods game, the amount of reward available was non-linear to reflect ecological settings. The more people who decided to produce, the less extra reward was available. In addition, we manipulated the relative cost of production in different summer and winter environments. We found that participants chose to produce more often

when the cost was lower but strikingly showed higher group benefitting behaviour than predicted by the Nash equilibrium in both summer and winter contexts. Together, these findings suggest that in ecological experimental settings, humans show high levels of prosociality and therefore when viewed from an ecological perspective people may be more willing to incur costs to benefit others than previously described.

Oral Presentations

O.01: Opportunity costs and opportunities lost

O.01.01 - Achieving Scale-Independent Reinforcement Learning Performance With Reward Range Normalization

Maeva Lhotellier [1], **Jeremy Perez** [2], **Stefano_ Palminteri_** [3]

[1] École Normale Supérieure-PSL University, [2] National Institute for Research in Computer Science and Automation, [3] École Normale Supérieure

Details

Study's objective

Experimental research suggests that people perceive rewards in a context-dependent manner. More specifically, they would compare new rewards to the maximum and minimum rewards they have received, a process referred to as range-adaptation (Bavard & Palminteri, 2023). Here we set out to explore what the advantage of performing range-adaptation might be from a computational point of view.

Methods

We conducted simulations comparing a standard reinforcement learning algorithm (delta rule) with a variant that incorporates range-adaptation. Using the two-armed bandit task paradigm, we systematically varied rewards magnitude from log10(−6) to log10(6) and measured models performance.

Results

First, as hypothesized, the results show that the performance of the standard (unbiased) model is scale dependent, i.e. its accuracy is high only for a narrow range of reward magnitudes. On the other hand, range-adaptation allows the algorithm to achieve higher performance regardless of the magnitude of the underlying rewards. In other words, range-adaptation achieves scale invariance.

Second, we show that these results can be explained by the inability of the standard model to deal efficiently with the exploration-exploitation dilemma. Indeed, the optimal value of the parameter controlling the degree of exploration depends on the magnitude of the rewards, such that a fixed

parameter value cannot be suitable for all magnitudes. Specifically, we show that reward range adaptation acts as a rational adjustment of the exploration parameter by bringing it close to its optimal value for any given magnitude.

We then extended our range-adaptation algorithm to tasks with deeper decision trees by integrating a temporal difference component. We designed a two-step, three-state task (with an entry state that could lead to either of two final states). Again, range-adaptation allowed for robust, scale-invariant performance. Conversely, standard model performance was above chance only over a small range of magnitudes.

Finally, we assessed the optimal values of the models' learning rates under different maximisation constraints. Range-adaptation outperformed the standard model over a wide range of parameters, and regardless of the variable optimized in the hyper-parameter search (instrumental vs. epistemic accuracy).

Conclusion

Our research provides normative ground for range-adaptation in the context of reinforcement learning. Processes previously viewed as sources of suboptimal decisions in laboratory settings may actually represent adaptive responses to the dynamic and open-ended nature of real-world rewards (Rustichini *et al*., 2023)

O.01.02 - Dynamic evolution of the decisional reference point across frontal brain regions in the monkey

Duc Nguyen [1], **Erin Rich** [2], **Joni Wallis** [3], **Kenway Louie** [1], **Paul Glimcher** [1]

[1] New York University, [2] Icahn School of Medicine at Mount Sinai, [3] University of California, Berkeley

Details

Objective: The reference point (RP) is a hidden benchmark against which a decision maker evaluates options. However, little is known about where the RP is represented in the brain and how it shapes the neural computations guiding the decision process. Here we reanalyzed a dataset from a previously published paper (Rich & Wallis, 2014) with two key features: (1) a stimulus-response task where earned within-trial reward value was fixed but cumulative across-trial reward (the RP) varied; and (2) neural recordings from multiple frontal brain areas involved in value-based decision-making. These features allowed us to investigate the neural representation of reference point and reference-dependent values.

Methods: Two NHP subjects were trained to perform an instructed response task. Each trial began with the onset of a reward bar indicating the current cumulative reward amount. Subjects were then presented with a visual cue instructing a left or right joystick response. Four cues were used, differing by correct response (left/ right) and valence (positive/ negative) in a 2×2 design. Positive cues resulted in one unit increase or no change in reward for correct and incorrect responses, respectively; negative cues resulted

in no change or one unit decrement in reward. Subjects received feedback through a change (or no change) in the reward bar size. After every sixth trial, subjects received a juice reward the magnitude of which was proportional to the size of current reward bar. In this task, the reward bar functions as the RP against which individual trial outcomes may be evaluated. Since individual trial outcomes were fixed but RP varied across trials, we examined the independent effects of reward and RP information on activity of 977 neurons recorded from six different frontal cortical regions (dACC, vACC, dlPFC, vlPFC, OFC, mOFC).

Results: Regression analyses showed a distributed representation of RP-related neural signals, with the involvement of different frontal brain regions at different points in the decision process. Early in the trial, a homogenous population coding of the RP was the strongest in the vACC. During cue onset, dACC showed a reference-dependent coding of cue valence. At the end of the trial, RP information modulated outcome coding in the dlPFC.

Conclusions: Our results reveal a distributed representation across frontal brain areas that correlates with the reference point and reference-dependent values. The evolution of these signals is mirrored by an anatomical transition from ventral to dorsal areas, providing a neurobiological substrate for the influence of the RP on decision-making.

O.01.03 - A computational model of regret

Joseph Heffner [1], **Robb Rutledge** [1]

[1] Yale University

Details

Regret is a negative emotion triggered by counterfactual thinking, a process of comparing reality to hypothetical alternatives (e.g., “Had I taken the umbrella, I wouldn’t have gotten wet”). Existing research focuses on anticipated regret (i.e., “How would you feel if X occurred”), identifying links to decision-making that may differ from those resulting from regret when actual outcomes experienced. To address this issue, we tested participants (N = 400) who engaged in a risky decision-making task while periodically reporting their momentary happiness. When participants opted for the safe option, we informed them of the counterfactual outcome (i.e., whether they would have won or lost if they had chosen the risky option).

We observed significant main effects of happiness being sensitive to both chosen gamble outcomes *(p*<0.001, *r*=0.75) and to counterfactual outcomes (*p*<0.001, *r*=0.26). A computational model that included a single parameter for counterfactual comparisons (real minus counterfactual outcomes) outperformed the standard happiness model without any parameters for counterfactual outcomes according to model comparison. This counterfactual parameter was positive on average (*p*=0.002), capturing the impact of counterfactual thinking on happiness. However, model comparison showed that this model was outperformed by a computational model including separate regret (counterfactual better than actual outcome) and relief (counterfactual worse than actual outcome) parameters. Interestingly, we found that the relief weights were, on average, significantly positive (*p*=0.003) while regret weights were not (*p*=0.4), suggesting considerable variation in the degree to which individuals are bothered by knowing that things could have gone better.

In summary, our findings provide evidence that counterfactual outcomes influence happiness. Although there's considerable variation in the extent to which individuals care about different types of counterfactual outcomes, our computational model shows that both relief and regret are critical determinants of monetary happiness.

Acknowledgements: J.H. and R.B.R. are supported by the National Institute of Mental Health (1R01MH124110).

O.01.04 - Sensitivity to distinct types of regret recruits separate striatal networks

Romain Durand-De Cuttoli [1], **Antonio Aubry** [1], **Long Li** [1], **Julian Sackey** [1], **Farzana Yasmin** [1], **Salma Elhassa** [1], **Sanjana Ahmed** [1], **Eric Nestler** [1], **Scott Russo** [1], **Brian Sweis** [1]

[1] Icahn School of Medicine at Mount Sinai

Details

Regret describes recognizing that an alternative action could have led to a better outcome. Recently, we discovered there may exist fundamentally distinct types of regret processed in separable circuits. These types are segregated by specific action-sequences tied to unique economic violations defined by whether an agent **(i)** forwent missed opportunities or **(ii)** made unfavorable investments.

Here, we characterized 40 mice on our neuroeconomic task "Restaurant Row." Mice had a limited time budget to forage for food investing in rewards of varying costs (delays, 1-30 s signaled by tone pitch) and subjective value (unique flavors). As previously published, regret trials were characterized by atypical economic violations where mice **(i)** rejected high-value opportunities only to be subsequently presented with worse offers or **(ii)** accepted low-value offers that usually in turn immediately evoke change-of-mind decisions. Importantly, we measured how each type of choice history is capable of altering future behavior as a readout of regret-related sensitivity. We extracted and leveraged whole brain tissue clearing to stain 275 regions for cFos neuronal expression, an activity-dependent biomarker of recent cell activity in order to capture a snapshot of brain-wide activation patterns.

We found that individual differences in sensitivity to each type of economic violation negatively correlated with one another. PCA analysis revealed 78.4% of variance was explained by PC1 with mice that were most sensitive to either scenario occupying opposite ends of this spectrum. We discovered a robust bidirectional change in cFos+ cell counts in the amygdala that strongly correlated with accumbens activation only in mice most sensitive to **(i)** missed opportunities. In contrast, hippocampus activation strongly correlated with accumbens activation only in mice most sensitive to **(ii)** change-of-mind decisions. In a separate cohort, we collected preliminary *in vivo* physiology recordings of large populations of single neurons in the amygdala and hippocampus as a first step toward understanding how decision algorithms employed by each structure integrate and converge, for instance, downstream in the accumbens in order to bind the value of hypothetical outcomes to specific unselected actions: a viable model for counterfactual credit assignment.

This work highlights how understanding separable aspects of regret processing, which we have shown can be differentially perturbed by stress and have ongoing efforts to characterize clinically in psychiatric

patients, can enhance the diagnostic nosology of circuit-specific computations mediating the perception and influence of one's prior actions.

O.02: Learning preferences

O.02.01 - Episodic memory is used to flexibly access features of past experience for decision making

Jonathan Nicholas [1], **Marcelo Gomes Mattar** [1]

[1] New York University

Details

Our decisions often require us to prioritize some features of our experiences over others. One way to do so is to focus on relevant features and discard those that are irrelevant. Yet learning which features to prioritize requires extensive experience. Moreover, features that are irrelevant now may become relevant in the future. These issues can be addressed by instead making decisions by sampling individual richly encoded experiences from episodic memory. Here we hypothesize that episodic memory is used to flexibly construct decision variables according to whichever features of the past are currently relevant for a choice. We test this hypothesis using a novel paradigm in which people were asked to make decisions about features that were present across multiple past experiences. Critically, making good choices in this task requires constructing a decision variable by summing the value of distinct episodes that share a common feature. Across two experiments (Exp 1: n=67; Exp 2: n=71), we find evidence that people use episodes to access features of past events on the fly during decision making. Participants' choices were better predicted by the value of subsequently recalled choice-relevant episodes relative to the veridical value of each choice (Exp 1: ELPD=16.48, SE=15.9; Exp 2: ELPD=17.78, SE=13.86; difference in cross validated expected log predictive density). Further, participants took longer to make choices that required referencing more episodes (Exp 1: β=0.05; 95% CI=[0.02,0.09]; Exp 2: β=0.06; 95% CI=[0.03,0.13]; main effect of the number of memories on decision response time). In a third experiment (Exp 3: n=50), we aimed to assess whether episodes are particularly useful when it is unclear which features should be prioritized ahead of time. To do so, we manipulated whether participants knew prior to encoding each episode which features would be needed for future choices. Only when this information was unknown did participants base choices on the value of individual episodes (Known: ELPD=-6.32, SE=10.19, Unknown: ELPD=13.14, SE=7.61) and take longer to make choices based on more episodes (Known: β=0.02; 95% CI=[-0.04,0.07], Unknown: β=0.08; 95% CI=[0.04,0.12]). Overall, these results suggest that episodic memory promotes adaptive choice when knowledge of multiple features is necessary.

O.02.02 - Drug and food craving in daily life

Sergej Grunevski [1], **Emmanuel Alvarez** [1], **Emma Schweitzer** [1], **Julia Kong** [1], **Anna Konova** [1]

[1] Rutgers University

Details

Craving—the intense, specific desire for something—is a common part of our everyday experience. Most people report experiencing craving of some kind, such as for palatable foods like chocolate. Craving for drugs is also a defining symptom of addiction and has been shown to predict drug use/relapse. In work aiming to model the shared, defining features of food and drug craving in the laboratory, craving has been shown to transform subjective value for the object of craving (and similar choice options) in a multiplicative and time-bound fashion. Here, we investigated the generalizability of this laboratory signature of craving by simultaneously surveying different types of cravings and craving moments in participants' daily lives. Treatment-engaged patients with opioid use disorder (OUD; *N*=67) and healthy controls (*N*=52) took part in a 28-day experience sampling study. Each day, participants were asked to report on their momentary desire for opioids and food (sweet/savory), their immediate willingness-to-pay (WTP) for opioids, a sweet, and a savory snack across different quantities, their current context, and past-hour exposure to drug cues (e.g., seeing heroin). Consistent with our study goal, we captured our participants' data across a wide diversity of real-world contexts, and some of our OUD participants' data in contexts known to provoke drug craving. Moments of drug desire were found to be separable from sweet or savory desire moments in OUD participants (repeated-measures *r*s<0.05, *p*s>0.15), even though participants who desired opioids also tended to desire sweets (Pearson's *r*=0.36, *p*<0.01). These desire moments were found to predict multiplicative shifts in subjective value specific to the desired object: drug WTP by drug desire in OUD participants (linear mixed model *b*=0.03, *p*<0.01) and sweet WTP by sweet desire in both samples (*b*=0.02, *p*<0.01). Lastly, we used hidden Markov models to test whether craving moments can be reliably and formally captured under a distinct craving "state," finding that our drug data could be well-characterized by a latent process where individuals transition between a "baseline" and a "drug craving" state with disproportionately higher drug value. Our findings show that individuals experience good-specific subjective value shifts in the real world where they can freely act on their environments while managing goal pursuits, social interactions, and affective states unafforded to them in the laboratory. Combining this sampling design with computational modeling may allow us to infer a behavioral read-out of distinct craving moments and types that could be targeted with just-in-time interventions for behavior change.

O.02.03 - Feeling Less, Risking More: The Role of Emotional Habituation in Risk-taking Escalation

Hadil Haj Ali [1], **Moshe Glickman** [1], **Tali Sharot** [1]

[1] University College London

Details

Objective: Anecdotally, extreme risk-taking can be traced back to minor risky behaviors that slowly snowballed over time. Here, we provide empirical evidence for a gradual escalation of risk-taking and reveal a mechanism supporting it**.** As individuals weigh risks, they can feel both negative emotions (anxiety) linked to potential costs and positive emotions (excitement) tied to anticipated rewards. We hypothesize that both reactions will decrease due to emotional habituation as individuals confront the same risky decisions. This should result in risk-taking escalation as (i) the negative emotion that initially curbs risk-taking evaporates, and (ii) decreased excitement will lead to larger risk-taking to reinstate the initial positive feelings.

Methods: We imitated physical risk by employing Virtual Reality (VR). On each of the 15 trials, participants (N=74) had a repeated opportunity to walk as far as they desired on a plank suspended at great heights. The magnitude of the virtual risk-taking was operationalized as the distance walked on the plank. GSR and self-reported anxiety and excitement provided indicators of the emotional response on each trial.

Results: First, our results confirm that VR is a valuable model for real-life risk-taking. In particular (i) subjects who took more risk in the VR task reported more risk-taking in real-life (β=0.3, t(70)=2.68, P=.009), and (ii) as subjects travelled further along the plank, they reported perceiving the situation as riskier (β=0.245, t(34.26)=3.77, P <.0001) and feeling more anxious (β=0.07, t(52)=2.01, P <.05). Second, we show that participants' risk-taking increased with repetition (β=0.3, t(70)=2.68, P=.009), while their emotional responses decreased, consistent with habituation (anxiety: β=-.04, t(72)=-3.3, P <.05; excitement: β=-0.09, t(72)=-6.1, P <.0001; GSR: β=-0.03, t(45)=-2.7, P <.001). Crucially, risk escalation was associated with emotional habituation; individuals who exhibited the fastest risk escalation (number of attempts needed to reach the plank's edge) exhibited the fastest rate of emotional habituation (the negative slope between time and emotional response; anxiety: R=0.42, P <.001; excitement: R=0.22, P =.039). A control study (N= 30) ruled out skill learning as a factor.

Conclusion: The results offer empirical evidence for the gradual escalation of risk-taking through repetition and tie the phenomenon to emotional habituation. Findings suggest that risk habituation can be reduced in unsafe environments (e.g., on the road) by restoring emotional reactions (e.g., by simulating a virtual accident).

O.02.04 - Bayesian modeling of action-induced preference changes

Guihua Yu [1], **Yaomin Jiang** [2], **Yunlu Yin** [3], **Lusha Zhu** [1]

[1] Peking University, [2] Max Planck Institute for Human Development, [3] Fudan University

Details

Objective: Accurately predicting everyday decision-making preferences is a cornerstone of consumer research. Classic economic theories suggest that actions merely reflect static preferences. However, behavioral and neuroimaging studies indicate that humans sometimes irrationally alter their preferences for certain things, places, or people after making decisions about them. It remains largely unknown, however, when, how, and why such action-induced preference changes occur. Here, we present a computational model in which preference shifts result from probabilistic perception of preferences: Internal evaluation of a stimulus emerges from a Bayesian inferential process that incorporates and justifies past decisions related to the stimulus.

Methods: The Bayesian Preference Change (BPC) model treats preferences as uncertain beliefs about the value of an item, and assumes these beliefs are updated via Bayesian inference based on the decisions people make regarding the item. Through model simulation, the BPC model not only replicates the widely reported effect of action-induced preference changes but also generates three novel, quantitative, and falsifiable predictions about factors modulating preference changes after a decision. We validated these predictions in laboratory experiments (N = 95) using an incentive-compatible task, where participants

indicated their real choice of items, and then rated their subjective preference for those items before and after their decisions.

Results: As predicted, the BPC model captured the phenomenon where accepting (rejecting) an item led to an increase (decrease) in preference ratings. Beyond explaining existing phenomena, we verified three model-derived predictions: larger preference increases after accepting an item: 1) when the picture was initially less preferred, 2) the uncertainty of the preference was larger, 3) or the acceptance was less frequent for the participant. Formal model comparisons further confirmed that the BPC model outperformed alternative models that either shifted mean preference values based on decisions or only considered the variability in revealed preferences.

Conclusions: Our findings demonstrate that preference changes can be quantitatively modeled as probabilistic inferences that rationalize past experiences. This Bayesian-based framework offers insights into future studies in neuroscience to test how past decisions are retrieved and reevaluated in the brain, guiding internal reward assessments for future decisions. It also provides actionable implications for predicting changes in consumer preferences after purchases or experiences, valuable for managers and practitioners.

O.03: Value representation in the brain

O.03.01 - Neural asymmetric representations predict the decoy effect

Asaf Madar [1], **Tom Zemer** [1], **Ido Tavor** [1], **Dino Levy** [1]

[1] Tel Aviv University

Details

The decoy attraction effect is one of the most well-known effects in behavioral economics. Over the last 40 years studies have replicated and investigated the effect, focusing mainly on structured stimuli with two explicit numerical properties. As a result, current computational models provide no explanatory or predictive tools for the occurrence of the attraction effect for natural or complex stimuli that cannot be represented by a few numerical properties. Here, we propose a novel method for determining which stimuli would elicit the decoy attraction effect and to what extent, without using any of their explicit properties, but rather based on similarity of neural representations between the options as measured using functional magnetic resonance imaging (fMRI). We use lotteries, which are known to elicit the decoy effect and have explicit numerical properties to test our approach, but we use only their neural representations to predict the occurrence of the attraction effect. First, subjects (n=26) inside the fMRI were presented with one lottery on the screen at a time and stated their willingness to pay to participate in the lottery using a Becker–DeGroot–Marschak (BDM) task. Then, for each subject we extracted the neural representation of each lottery and computed the neural similarity between each pair of lotteries. A different sample (n=62) performed a standard behavioral decoy effect choice task outside of the fMRI, choosing between binary or ternary sets of lotteries that were presented to the first group inside the fMRI. Based on this behavioral experiment, we computed the magnitude of the decoy effect for each set of lotteries, which ranged from -9% to 24% change in preference for the target option when comparing between the binary and ternary sets. Finally, we used the average neural similarity between the lotteries from the fMRI sample to predict the magnitude of the decoy effect that each set of lotteries elicited in the

behavioral sample. We found significant predictions of the decoy effects (R^2_adj=0.7086, p=0.0044) using neural representation similarities. We further find that the patterns of representational similarity that predict the decoy effect are in line with the known patterns predicted by current models using the known stimuli property space (amount and probability). Thus, we show for the first time, that the decoy effect can be predicted without relying on any explicit properties of the stimuli. Our method can therefore generalize the decoy effect to novel stimuli, absent of explicit properties, and provide new insights for the neural mechanisms of decision-making biases.

O.03.02 - Flexible hippocampal representation of abstract boundaries supports memory-guided choice

Mariachiara Esposito*[1], Lubna Abdul*[1], Marta Rodriguez Aramendía[1], Raphael Kaplan[1], Ameeruddin Ghouse[1] - [1] Universitat Jaume I, Castelló de la Plana, Spain

[1] Universitat Jaume I

Details

Cognitive maps in the hippocampus encode the relative locations of spatial cues in an environment and dynamically adapt their representation when boundaries geometrically change. In parallel, cognitive maps can represent abstract knowledge, yet it's unclear whether the hippocampus is sensitive to geometric changes to the borders, extreme coordinates, of abstract knowledge spaces. Addressing whether abstract boundaries are processed by the human hippocampal-prefrontal circuit in a similar way as physical borders, we developed a two alternative forced choice(2AFC) fMRI task.
Here, participants made similarity judgements on either the price or freshness of various fruits and vegetables. Unbeknownst to participants (N=29), the continuous price and freshness variables for each product formed two-dimensional(2D) abstract spaces, with four of the goods lying at the extreme coordinates(boundary goods), one good fixed near the center(landmark good), and sixteen cued goods that were located within the boundaries. Despite being unnecessary to accurately make decisions, participants conserved a 2D map-like representation of abstract boundaries during a surprise drag-and-rate task after the fMRi paradigm, where the precision of their representation related to prior choice accuracy. Finding that the hippocampus and mPFC represent the Euclidean distance of a decision cue to the most proximal boundary during decision making, we then tested whether hippocampus and mPFC signals were sensitive to abstract boundaries. Conducting MVPA decoding analyses on hippocampal and mPFC patterns elicited by boundary goods, we found that boundary-defined contextual identity can be accurately decoded from hippocampal fMRI signals. Relating classifier accuracy to task behavior, we then asked whether hippocampal classifier accuracy related to participant choice accuracy. Classifier accuracy didn't correlate with general task performance. However, since the hippocampal classifier appeared to be biased towards one shaped context versus another, we then investigated whether the classified bias towards the favored shape corresponded with participants' performance biases on the 2AFC task. Relating the difference in task performance in the different shaped contexts to the hippocampal classifier bias, there was indeed a significant correlation with classifier accuracy. Consequently, we provide evidence that the human hippocampus flexibly represents abstract knowledge boundaries and that this process supports memory-guided choice behavior.

O.03.03 - Dissociable Neural Representation for Gains and Losses Under Uncertainty

Ohad Dan [1], **Ifat Levy** [2]

[1] Yale University, [2] Yale School of Medicine

Details

Background: Ample evidence demonstrates that behaviorally, losses loom larger than gains. What is the neural basis of this asymmetry? A common neuroeconomic framework to explain value representation is the common currency hypothesis, which postulates that the brain encodes value, across domains, using a single centralized set of areas, including the ventral striatum and ventromedial prefrontal cortex (vmPFC). According to this theoretical framework, both positive and negative outcomes, 'gains' and 'losses', should be represented by these same brain areas, with neural activation levels monotonically ranging from great losses to little losses, to small gains, to great gains. While some seminal previous work is consistent with this framework, little research directly tested this hypothesis.

Objective: Test whether subjective value of gains and losses is represented by dissociable brain mechanisms.

Methods: Using fMRI, we recorded the neural activity of 120 healthy participants as they performed a decision-making task. Participants made a series of binary choices between certain $5 outcomes and uncertain outcomes ranging from $6 to $120. The uncertain outcomes were presented with known probability ('risk') or partially unknown probability ('ambiguity'). Importantly, half of the trials were framed as gains (e.g., a choice between winning a certain $5, and a 50% chance of winning $12) and the other half as losses (e.g., a choice between a certain loss of $5, and a 50% chance of losing $12). To identify the neural signature of subjective value, we fit participants' choices to a standard economic model and infer individual risk and ambiguity attitudes. Using these inferred attitudes, we extrapolated post hoc the subjective values participants attributed to different lotteries and searched for neural clusters whose activity correlated with subjective value.

Results: Observing brain areas whose activity correlated with each domain, we found distinct activation maps for gains and losses. Consistent with previous literature, we found that activation levels in the vmPFC correlated with the subjective value in the gains domain. In contrast, subjective value in loss trials was represented by more lateral areas such as the dorsolateral prefrontal cortex.

Conclusions: We find an intriguing deviation from the common currency hypothesis. Our results suggest that rather than a single valuation system, the brain employs distinct neural pathways depending on whether the represented values are framed as gains or losses. This nuanced understanding of the brain's value computations sheds new light on our understanding of value-based decision-making.

O.03.04 - Reproducible neural activity during delay discounting across normative and psychiatric samples

Lauren Patrick [1], **Min Su Kang** [1], **Tinashe Tapera** [1], **Kosha Ruparel** [1], **Mark Elliott** [1], **Sage Rush-Goebel** [1], **Anna Xu** [1], **Matthew Cieslak** [1], **Caryn Lerman** [1], **Theodore Satterthwaite** [1], **Daniel Wolf** [1], **Joe Kable** [1]

[1] University of Pennsylvania

Details

Intertemporal choices, choices between smaller sooner and larger later rewards, are pervasive in everyday life. During intertemporal choice, people tend to discount, or devalue, future rewards. Decades of research in decision neuroscience have examined the neural mechanisms underlying intertemporal choices and delay discounting. Understanding these neural processes is important in part because they are a promising measure for transdiagnostic research. Because a central challenge to brain-behavior phenotyping is reproducibility, our work aimed to establish recommendations for analyzing delay discounting neuroimaging data in a reproducible manner. To this end, we analyzed two independent datasets (Study 1 N=154 and Study 2 N = 158) that used a similar delay discounting paradigm in either healthy adults or individuals with mixed psychiatric diagnoses. We systematically evaluated the reproducibility of common analytic methods used in neuroimaging studies of delay discounting across datasets. We examined the main effects of commonly used contrasts in the literature (task vs. rest, choice, subjective value, subjective value difference, chosen value, objective amount, length of delay, reaction time), as well as their associations with individual differences in discounting behavior. Our findings suggest that computationally derived variables (e.g., subjective value) outperformed objective variables (e.g., amount, delay) in explaining brain activity. Moreover, our analyses revealed that activity in ventral striatum and ventromedial prefrontal cortex is best described by chosen value and a separate region of medial prefrontal cortex tracks subjective value difference. Additionally, we conclude that individual differences in discounting modulate choice and value-related activity, rather than general task-related activation.

O.04: Attention and efficiency

O.04.01 - A model of goal-directed search in multi-attribute decision making

Sebastian Gluth [1], **Jörg Rieskamp** [2]

[1] University of Hamburg, [2] University of Basel

Details

Objective. In an increasingly complex world, decisions problems are often characterized by the presence of many choice options that are described by several attributes. Thus, humans must strike a balance between the desire to make informed decisions and the need to limit invested resources. Here, we propose that humans achieve this balance by searching for and thus devoting attention to relevant information in an efficient and goal-directed, but not strictly optimal way. Thereto, we developed a hierarchical Bayesian cognitive model of information search in multi-attribute decisions, in which the search is assumed to be governed by the desire to quickly identify the option that meets the current choice goal best. At the core of our theory is a myopic search rule that looks one step ahead and prefers sampling more promising choice candidates.

Methods. We tested our theory in a preregistered eye-tracking experiment, in which we minimized the influence of bottom-up attention on visual search, thus allowing us to test a series of top-down attention effects that our theory predicts. N = 65 participants first rated sets of 30 smartphones and 30 hotels based on 3 attributes each (e.g., size, battery, and memory of smartphones) and then chose between smartphones and between hotels (120 trials each). Importantly, attributes were presented as isoluminant icons of similar size that slowly changed their position on the screen to cancel out the well-known bottom-up attention effects of position, brightness, and size.

Results. Our design allowed us to reproduce many existing findings on the interplay of attention on choice. Participants were more likely to choose options they fixated more often ($p < .001$), were more likely to look at more attractive options ($p < .001$), and were more likely to first look at the most important attribute ($p < .001$). In hotel choices, the tendency to exhibit a more alternative-wise search pattern was correlated with more equal weighting of attributes ($p = .014$). Critically, our model predicts all these effects in the current binary choice set, and additional simulations indicate that these predictions generalize multi-alternative decisions.

Conclusions. Our work offers a principled framework for understanding and predicting search and choice patterns in multi-attribute decisions, going beyond existing work that only use attention data to inform behavioral model predictions. At the same time, our theory can accommodate seemingly irrational search patterns in binary decisions that pose a challenge for extant models of rational inattention.

O.04.02 - Assessing Uncertainty in Valuation Through Direct Behavioral Measures

Hui-Kuan Chung [1], **Nick Doren** [2], **Hsin-Hung Li** [3], **Philippe Tobler** [1]

[1] University of Zürich, [2] University of Zurich, Zurich, [3] Ohio State University

Details

Decision-making relies on stochastically and noisily represented values. In perception and number processing, it is well established that noise increases with intensity of the input. This phenomenon may explain individual variations in risk aversion. Typically, these approaches infer uncertainty from trial-to-trial variation in perceptual judgments. However, such variation can have multiple causes and more direct behavioral measurement of uncertainty in valuation is needed. We characterized such a measure and investigated its neuropharmacological underpinning. We used a modified Becker–DeGroot–Marschak approach (Wang et al., 2007; Dost & Wilken, 2012) to assess uncertainty in valuation and risk tolerance. Instead of reporting a single point willingness-to-pay (WTP) for a lottery, participants indicated symmetric minimum and maximum WTPs for each lottery. The reported range constitutes a direct behavioral index of their uncertainty in valuation, whereas the mean of the range corresponds to the point WTP. We operationalize state risk aversion in a trial-by-trial fashion as difference between the expected value (EV) of the lottery and the mean WTP, with a higher number indicating higher risk tolerance. In Experiment 1 (n=26), we found that uncertainty in valuation (range) increased with EV (β = 0.25, 95% CI range = [0.15 0.35]). Additionally, confidence rating, assessed separately in different runs, decreased with uncertainty in valuation (β = -0.09, 95% CI range = [-0.16 -0.01]). Strikingly, trial-by-trial uncertainties (ranges) were not related to trial-by-trial risk aversion (β = -0.17, 95% CI range = [-0.44

0.09]). In Experiment 2, we used three psychoactive substances – noradrenergic reuptake inhibitor reboxetine, the dopamine reuptake inhibitor methylphenidate (MPH), and the cholinergic receptor agonist nicotine, in a randomized, placebo-controlled, double-blind, and incentive-compatible design. We hypothesized these neuromodulators to reduce noise compared to the placebo group. 160 (80 women) healthy volunteers (mean age ± SD = 23.63 ± 3.65 years) were assigned to either placebo or a psychoactive drug group, with 40 participants per group. First, we replicated that uncertainty (range) increased with EV. Second, the MPH group showed a reduction in valuation uncertainty compared to the placebo group (β_{MPH}-$\beta_{placebo}$ = -6.69, 95% CI range = [-12.85 -0.61]). However, again, the trial-by-trial uncertainty (ranges) were unrelated to trial-by-trial risk aversion (β = -0.02, 95% CI range = [-0.1 0.05]). These data indicate that dopamine increases the precision of value representations but that this precision is not directly related to risk aversion.

O.04.03 - Do glitches in the OFC neural code explain irrational choices? A neuro-computational approach to value synthesis.

Juliette Bénon [1], **Jean Daunizeau** [1]

[1] Paris Brain Institute

Details

The orbitofrontal cortex (OFC) is known to play a key role in integrating environmental features and signaling value. However, the functional contribution of OFC neurons to value-related computations is still unclear. For example, the OFC may be operating the comparison of options' values for decision making. But another possibility is that OFC neurons are integrating value-relevant features into value estimates, leaving their comparison for downstream decision circuits. This distinction is important, because imperfections in the underlying neural system architectures would yield different forms of irrational behavior. In this study, we aim at understanding how biological constraints in these neural architectures could affect decision-making.

To do so, we built models of both neural activity and behavior. We trained recurrent neural networks (RNNs) to perform either value synthesis or value comparison, under the constraint that options' features are sampled progressively in time. For both scenarii, we considered different value coding frames, in terms of options spatial position (left/right), temporal order (first/second) or attentional focus (attended/unattended). We then compared the models' neural signatures to electrophysiological data recorded by Hunt et al. (2018) in the OFC of two macaque monkeys during a binary decision task.

We found that key OFC representational geometry properties could emerge under both functional scenarii, but only in specific value coding frames. Furthermore, when disrupting the RNNs' feedback connections to explain monkeys' irrational choices, the representational geometry of RNNs becomes even more similar to that of OFC neurons. Interestingly, the ensuing sparsity of neural representations increases, while the amount of energy used by the units decreases. The RNN also becomes better at maintaining optimal choices when it is lesioned.

These results suggest that apparently non-optimal decisions might actually be the best balance when considering not only the reward obtained through the behavior, but also biological constraints on the OFC neural code such as a limited energy budget and the need for robustness.

O.04.04 - Exploring attentional mechanisms underlying value normalization in human reinforcement learning

Romane Cecchi [1], **Stefano_ Palminteri_** [1]

[1] École Normale Supérieure

Details

Objectives

The choices humans make are heavily influenced by context, with the perceived value of an option being strongly shaped by other alternatives. This contextual valuation is evident in reinforcement learning (RL), where values are learned through trial and error. One proposed explanation is range adaptation, where options are rescaled based on the range of values in a given context. Moreover, a recent study introduced a weighted range normalization model to explain why, when choosing between three options with equally distributed values, the mid- and low-value options appear subjectively closer (Bavard & Palminteri, 2023). Although this model is descriptively accurate, it fails to explain its cognitive underpinnings. We hypothesized that attentional mechanisms may play a crucial role in shaping the subjective valuation of options in RL tasks.

Methods & Results

To test this, we conducted three eye-tracking experiments, each with 35 participants engaged in a choice task within an RL framework. Participants chose between three non-overlapping, evenly distributed value options with full feedback in a learning phase, followed by a transfer test in which they chose between all possible pairs without feedback.

In the first experiment, we manipulated attention by preventing the selection of high-value options in certain trials, thus biasing choices towards mid-value options. This corrected the subjective valuation of mid-value options in the transfer test, as they were chosen more often than low-value options ($t_{34} = -9.5$, $p = 5.10^{-11}$). Experiments 2 and 3 further investigated the role of attention by manipulating visual attention independently of choice. We found that the subjective valuation of mid-value options in the transfer test was corrected by increasing the salience of their stimuli ($t_{34} = -3.7$, $p = 8.10^{-4}$), but not their outcomes ($t_{34} = -1.7$, $p = 0.108$). These results suggest that attentional allocation, specifically to option stimuli, influences the subjective valuation of options.

To account for these findings, we developed an RL model incorporating attentional mechanisms, where the time spent looking at options before choice directly affects their absolute value before range normalization. This attentional range model performed comparably (BIC E1: weighted=346; attentional=362) or better than (BIC E2/3: weighted=322/266; attentional=309/258) the weighted range model, highlighting the importance of attention in shaping value representations.

Conclusion
Overall, our results provide insight into how attention interacts with value updating in RL and shed light on the cognitive mechanisms underlying context-sensitive valuation in human decision-making.

O.05: Decision making in complex & social contexts

O.05.01 - The earning-spending behaviour and the underlying neural mechanism in macaque monkeys

Hualei Wang [1], **Hangyu Si** [1], **Tianming Yang** [1]

[1] Chinese Academy of Sciences

Details

Objective: Earning and spending are two fundamental behaviours in our daily economic lives. Yet the mechanisms by which the brain makes choices regarding earning and spending remain largely unexplored. **Methods:** We trained three macaque monkeys to perform a novel task where they could freely choose between earning virtual tokens and spending them for juice rewards by fixating on two targets, representing earning and spending, respectively. Tokens earned went into a virtual wallet, and the monkeys received juice rewards only when they chose to spend them. The rate of earning and spending was indicated by target shapes, which we varied across trials. To understand the neural basis of this earning-spending behaviour, we recorded single-unit activity from the anterior cingulate cortex (ACC, n=278), the orbitofrontal cortex (OFC, n=456), and the dorsal lateral prefrontal cortex (DLPFC, n=102). **Results:** (1) We observed that the monkeys' preference between earning and spending shifted according to the values of shapes. Notably, this preference was additionally biased by the monkeys' current token savings, with a higher tendency to spend when more tokens were in their wallet. (2) While all three brain areas encoded key value information involved in the task, only the OFC neuronal activity reflected the earning-spending balance modulated by current token saving. **Conclusions:** Our study provides evidence that monkeys can understand sophisticated economic concepts of earning, spending, and savings in a token-based economy, and exhibit human-like decision-making patterns. The results demonstrate the important role of the prefrontal cortex, especially the OFC, in earning-spending decisions.

O.05.02 - Cognitive mechanism of decision-making under social responsibility

Dongqi (Sherry) Bao [1], **Micah Edelson** [1], **Todd Hare** [1]

[1] University of Zürich

Details

Objective: People often need to take responsibility for others, with widespread and lasting impacts on both themselves and those affected. Responsibility for others has been shown to alter risky decision-making and delegation behaviors. However, the cognitive mechanisms behind this altered behavior remain underexplored.

Method: We conducted a series of experiments in humans (N=344) using perceptual decision tasks in which the decision maker did or did not have responsibility for others. Across experiments, we vary the availability of decision evidence, presence of feedback, and social context. By measuring choice certainty, accuracy, and response time, we aim to delineate the cognitive processes affected by social responsibility. Furthermore, we propose and empirically test a normative framework to explain responsibility aversion.

Results: Participants showed equal accuracy but lower confidence (limited stimulus display $p<0.001$;unlimited display $p=0.02$) and longer response times ($p<0.001$;$p=0.05$) when responsible for others. Furthermore, decreases in confidence are associated with slower response times($r=-0.47$;$r=-0.46$), but not accuracy. Feedback enhanced people's confidence in general, but didn't close the confidence gap created by responsibility for others. Additionally, the lower confidence with responsibility cannot be explained by simply having higher total stakes in the decisions.

Using a DDM, we find that the boundary is higher when taking responsibility for others. This suggests a lower subjective cost of sampling information and/or a higher certainty threshold. A computational model of the metacognitive process revealed that the change in metacognitive bias parameters mainly explained the change in confidence between conditions, instead of metacognitive noise(efficiency).

When given the option to delegate decisions to experts (with accuracy of 70%/90%), the majority exhibited responsibility aversion with high-accuracy experts($p<0.001$). Our normative model predicted that lower confidence when forced to take responsibility for others is related to responsibility aversion, which is confirmed in the high ($p=0.01$), but not low-accuracy expert condition. This difference can be attributed to a higher tendency to make decisions independently in the low-accuracy expert condition. Meanwhile, We find a negative correlation between delegation bias and total task accuracy ($r=-0.25$), which is predicted by our model.

Conclusions: Our findings reveal that taking responsibility for others results in alterations to the evidence accumulation and metacognitive process, which in turn contributes to responsibility aversion, mediated by confidence in the delegate.

O.05.03 - Do we advise as one likes? The alignment bias in social advice giving

Xitong Luo [1], **Lei Zhang** [2], **Yafeng Pan** [1]

[1] Zhejiang University, [2] University of Birmingham

Details

(1) Study's objective:

Advice giving provides a valuable source of knowledge and information during social interaction. We give advice to influence the advisees, but is our advice, in turn, also influenced by the advisees? For instance, despite individual advisors may personally prefer black ties, they might find themselves recommending red ones to accommodate a friend's preconceived preference for red ties, leading to an *alignment bias* in advice giving. However, very little is known about this alignment bias during advice

giving, including its underlying motivations and how it impacts the advice quality, and even advisors' future beliefs. Here, we addressed this nuanced phenomenon of the alignment bias in social advice giving for the first time. Leveraging sophisticated experiments and reinforcement learning (RL) models, we offer a comprehensive cognitive computational account of (1) how much advice giving is biasedly aligned with advisees, (2) how this alignment bias emerges, (3) what are the consequences of such alignment bias during advice giving.

(2) Methods used:

We conducted three closely connected experiments (N = 284) to systematically address these questions. Across all three experiments, participants were instructed to provide advice either in isolation or after observing opinions from an advisee. The advisees' preferences for taking advice were either concealed from the advisors (Study 1) or disclosed to advisors through feedback presentation (Study 2 and Study 3). We additionally manipulated the advisees' overall acceptance rate to induce the perception of interacting with either mild (Study 2) or stern (Study 3) advisees.

(3) Results obtained:

First, we robustly observed that advisors tended to align their advice with advisees' opinions (i.e., alignment bias). Leveraging RL models, we then unveiled that advisors' motives (opinion conformity vs. preference conformity) to ingratiate with advisees underlay this alignment bias. Finally, the alignment bias impaired both the quality of advice when interacting with advisees with lower performance, and a compromise in the impartialness of advisors' beliefs.

(4) Conclusions:

These findings highlight a counterintuitive disposition of advisors — susceptibility to social approval— that may lead to the dissemination and exacerbation of misinformation within social encounters. It also demonstrates a striking but reasonable reality for the bi-directional nature of influence transmission: advisors, who are typically perceived as impartial and in control of influencing others, can also be swayed by social acceptance and influenced by advisees, who are conventionally seen as the ones being influenced.

O.05.04 - Individual differences in psychopathy are associated with increased social information use in a donation task

Iris Vilares [1], **Tao Jin** [2]

[1] University of Minnesota, Twin Cities, [2] University of Minnesota

Details

Introduction: Prosocial behaviors, actions intended to help or benefit others, are susceptible to peer influence. Previous research on prosocial conformity mainly focused on how others' mean behavior affected individual behavior. However, how the variance in social information affects individual prosocial

behavior remains poorly understood. Moreover, such susceptibility to others' decisions varies largely across individuals.

Objective: Here, we asked whether the mean and variance of social information affected prosocial behavior, and if individual differences in psychopathology influenced these effects.

Methods: In two pre-registered experiments (discovery sample: N=353, replication sample: N=371), participants completed a charitable donation task at first. After observing five others' donations, they had the chance to change initial donations. We manipulated both the mean and variance of social information: participants observed generous/ stingy donations with low/ high variance. We also built reinforcement learning models that analyzed the extent to which individuals learn others' generosity and integrate social information into their own donation decisions.

Results: We found that people observing generous donations from others increased their average donations, whereas those observing stingy donations decreased their own donations (Sample1: $F(1, 349)=263.81$, $p<.001$; Sample 2: $F(1, 367)=270.48$, $p<.001$). Moreover, the dispersion of individual donations from the group norm decreased more upon observing more consistent others' donations compared to more varied contributions (Sample1: $F(1, 349)=3.96$, $p= 047$; Sample 2: $F(1, 367)=18.60$, $p<.001$). In terms of individual differences, psychopathic traits (from the Levenson's Self-Report Psychopathy Scale) were positively correlated with changes in donation amounts after observing social information (Sample1: $r=0.15$, $p=0.004$; Sample2: $r=.21$, $p<.001$). Computational models including both an initial willingness to donate to a particular charity and a weight on observed social information revealed that people with higher levels of psychopathic traits were more susceptible to social information by placing higher weights on social information when making their second donation decisions (Sample1: $r=.19$, $p=0.004$; Sample2: $r=.25$, $p<.001$).

Conclusions: Our findings, replicated in two samples, suggest that the average and variability of social information impacted the level and clustering of individual donations respectively, and psychopathy may be associated with higher use of external social cues to make prosocial decisions.

Poster Spotlights

Poster Spotlights I

PS.01.01 - Sex differences in change-of-mind neuroeconomic decision-making is modulated by LINC00473 in medial prefrontal cortex

Romain Durand-De Cuttoli [1], **Orna Issler** [2], **Benjamin Yakubov** [3], **Nusrat Jahan** [3], **Aisha Abid** [3], **Susanna Kasparov** [3], **Kastalia Granizo** [3], **Sanjana Ahmed** [3], **Scott Russo** [1], **Eric Nestler** [1], **Brian Sweis** [1]

[1] Icahn School of Medicine at Mount Sinai, [2] New York University, [3] Hunter College

Details

Ruminating on the past when deciding to change one's mind is a complex cognitive phenomenon involving a continuous re-appraisal of the trade-off between past costs and future value. Recent work modeling this behavior across species established contributions of alterations in this decision-making process to psychopathology. The incidence of depression is two times higher in women than in men and can manifest different symptomatology, including negative rumination. Strides made in neuroeconomics have segregated stages of a decision stream into its component parts, isolating properties of change-of-mind re-evaluative decisions, and pointing to the medial prefrontal cortex (mPFC) as a critical hub for regulating reward value, self-control, and mood. We recently demonstrated that mPFC activity was causally linked to change-of-mind decisions. Recent transcriptomic analyses of human postmortem mPFC tissue obtained from individuals diagnosed with depression revealed a female-specific decrease in the long intergenic non-coding RNA: LINC00473 expression in the mPFC. We investigated in mice the actions of mPFC LINC00473, whose role in cognition has not yet been studied. While LINC00473 is not normally expressed in rodents – it arose in the primate lineage – expressing LINC00473 in mouse mPFC neurons demonstrated pro-resilient effects in females, but not males, using simple stress assays. Here, we characterized complex decision-making behavior in male and female mice longitudinally in our neuroeconomic foraging paradigm, Restaurant Row, following virus-mediated LINC00473 expression in mPFC neurons. On this task, mice foraged for their primary source of food among varying costs (delays) and subjective value (flavors) while on a limited time-budget during which decisions to accept and wait for rewards were separated into discrete stages of primary commitments and secondary re-evaluations. We discovered important sex differences in decision-making behavior. LINC00473 expression selectively influenced multiple features of re-evaluative choices, without affecting primary decisions, in female mice only. These behavioral effects included changing how mice (i) cached the value of the passage of time and (ii) weighed their history of economically disadvantageous choices. Both processes were uniquely linked to change-of-mind decisions and underlie the computational bases of distinct aspects of regret-related counterfactual thinking, potentially restoring neuroeconomic vulnerabilities in females. These findings reveal a key bridge between a molecular driver of stress resilience and psychological mechanisms underlying sex-specific decision-making proclivities.

PS.01.02 - Probability distortions reflect boundary effects under cognitive noise

Saurabh Bedi [1], **Gilles De Hollander** [1], **Christian Ruff** [1]

[1] University of Zürich

Details

Experiments consistently reveal deviations from classical economic theory. One notable example is probability distortions in experiments involving risk. Intriguingly, apparently similar distortions can also be observed in simple perceptual judgments of relative frequency, and more recently in experiments involving “complexity” but not risk. We propose a comprehensive mechanistic explanation for this phenomenon based on a domain-general model of resource-rational cognitive representation of bounded variables. Our model characterizes probability distortions observed in perception, risk, and complexity experiments as central tendency effects within the naturally bounded range of probabilities (0-1), which emerge mechanistically from cognitive noise. Unlike existing Bayesian and efficient coding models, our approach doesn’t assume specific prior distributions. We show that this parsimonious model captures both the distortion and variance in responses seen across experiments, and we directly test the causal validity of our model in an experiment on probability distortions in risky choice.

In our experiment, participants (N=63) saw lotteries with varied probabilities but fixed payoffs and reported corresponding certainty equivalents. We used a within-subject 3x2 factorial design with three probability ranges (naturally bounded from 0-1, and artificially bounded from 0-0.5 and 0.5-1) and two cognitive noise levels (complexity of presented probability information).

Our results align with all our model-predicted preregistered hypotheses. Specifically, we observed overestimation of small probabilities (natural boundary at 0) and underestimation of large probabilities (natural boundary at 1) that amplified with increased cognitive noise ($p<2e-16$). Additionally, we found comparable over/underestimation above/below the artificially induced boundary at 0.5 in the high cognitive noise condition ($p<2e-16$) and an interaction effect of the artificial boundary at 0.5 and cognitive noise ($p<2e-16$). Finally, we observed decreased variance in valuation close to the artificial boundary at 0.5 for high cognitive noise ($p<0.001$). This confirms our central model assumptions: The characteristic inverse S-shape pattern of probability distortions causally emerges from boundary effects and cognitive noise.

Our theory predicts distortions and variance in both economic and perceptual experiments involving probabilities. However, we show with model simulations that it is applicable also to other puzzling choice phenomena, such as stake size effects and overall value effects on variability/accuracy. These model predictions can be tested with causal manipulations in future experiments.

PS.01.03 - Serotonin increases willingness to wait for delayed rewards in humans

Karolina Lempert [1], **Laura Zaneski** [2], **Arjun Ramakrishnan** [3], **Daniel Wolf** [2], **Joe Kable** [2]

[1] Adelphi University, [2] University of Pennsylvania, [3] Indian Institute of Technology Kanpur

Details

Objective: People often have to decide how long to wait for rewards that will arrive at an uncertain time in the future. Previous work has suggested that people approach these willingness-to-wait decisions by trading off the benefits of continued waiting against opportunity costs of waiting, the balance of which can change as they wait. Thus, willingness-to-wait decisions might share neural mechanisms with foraging decisions, which also involve an ongoing comparison between a current resource and possible alternatives. Dopamine and serotonin have been proposed to play opposing roles in foraging, with

dopamine promoting exploration and serotonin promoting exploitation. Here we investigated their roles in willingness to wait, in a pre-registered, crossover, double-blind, placebo-controlled study in young adults. We examined how increasing dopamine (with 200 mg levodopa) and increasing serotonin (with 20 mg escitalopram) affected willingness to wait. Based on their proposed roles in foraging, we expected that levodopa would decrease willingness to wait and escitalopram would increase it. Methods: Participants (n = 42) completed three sessions, each one week apart. At each session, they received either placebo, escitalopram, or levodopa (counterbalanced order), and then performed decision-making tasks, including a willingness-to-wait task. On each trial of this 10-minute task, they waited through a random delay (0.2-40 s) for a 10¢ reward. They did not know how long the 10¢ would take to arrive, and they could quit waiting at any time and go to a new trial. Reward delays were distributed such that there was no optimal time at which to give up waiting. We plotted the probability that participants were still waiting for the reward at any given time point within the 40 s delay period; the area under this curve (AUC) was our measure of willingness to wait, with higher values indicating longer wait times before quitting. Results: We first decided on the inclusion of covariates according to our pre-registered plan, before running two mixed-effects regressions (one for levodopa, one for escitalopram) to test for the effects of drug on AUC. After controlling for anxiety levels and condition order, there was a significant effect of taking escitalopram on AUC, such that people were more willing to wait for delayed rewards after taking the serotonergic drug (b = 3.93; p = 0.012). There was no effect of levodopa on willingness to wait (b = -0.64; p = 0.398), however. Conclusions: This result nicely complements the rodent literature, which has shown a critical role for serotonin in waiting for rewards that are delayed by seconds. Possible mechanisms will be discussed.

Poster Spotlights II

PS.02.01 - Age-dependent changes in hippocampal contributions to decision-making

Michael Cohen [1], **Karolina Lempert** [2], **Kameron Macnear** [3], **Frances Reckers** [4], **Laura Zaneski** [4], **David Wolk** [5], **Joe Kable** [4], **Camilla Van Geen** [4]

[1] University of Chicago, [2] Adelphi University, [3] University of Illinois - Urbana-Champaign, [4] University of Pennsylvania, [5] Hospital of the University of Pennsylvania

Details

This study investigates how changes in hippocampal-dependent processes during healthy aging affect different forms of decision-making. A sample of 86 participants (45 younger and 39 older adults) first underwent fMRI while they performed a choice task in which accurate performance relied on retrieving value information from a single past episode. The same group of people were then also scanned during a reinforcement learning (RL) task in which they learned the average value of four stimuli incrementally across repeated exposures. The traditional view of multiple memory systems posits that these two paradigms should rely on distinct neural substrates, with episodic memory in the first task recruiting the hippocampus and RL in the second task evoking neural activity in the striatum and ventromedial prefrontal cortex (O'Doherty et al., 2003; Squire & Zola, 1996). Given that we can simultaneously characterize age-related neural changes in either system, our experiment is well-positioned to test the validity of this distinction: we compare group-level performance on each of the two tasks, as well as the neural activity that underlies it, in order to measure whether age-related cognitive impairments preferentially affect one function or the other. We find that older adults were especially impaired at making decisions from

episodic memory (age*reward interaction: b = 0.41, p = 0.016), and showed reduced activity in hippocampus and medial temporal lobe at encoding. In the reinforcement learning task, older adults accurately learned the value of the four stimuli (age*learning block interaction: b = 0.004, p = 0.84), and showed no differences in reward prediction error signaling in the striatum. Nonetheless, older adults did show subtle performance deficits in RL, especially at the end of learning (age difference in proportion correct choice on the final block: t = -1.86; p = 0.06). Using representational similarity analysis to assess how the hippocampus tracks distinct stimulus identities, we can link these subtle behavioral impairments to disrupted patterns of stimulus-specific neural activity. Together, these findings show that changes in hippocampal processing with healthy aging underlie both larger impairments in decisions based on episodic memory and more subtle impairments in incremental learning.

PS.02.02 - Creative choices rely on subjective values and suffer from cognitive biases

Sarah Moreno-Rodriguez [1], **Alizée Lopez-Persem** [1], **Emmanuelle VOLLE** [2]

[1] Paris Brain Institute, [2] Sorbonne University

Details

Creativity is defined as the ability to produce ideas that are both original and adequate. According to the dual-process theory, creativity involves a generation phase, during which individuals produce candidate ideas, and an evaluation phase, during which individuals assess and select the best idea among the candidates. Previous research (Moreno-Rodriguez et al., *in prep*) indicates that this evaluation phase features the subjective valuation of candidate ideas via the human reward system, also known as the Brain Valuation System (BVS). The current study's goal was to further investigate the decision-making mechanisms involved in creative thinking by focusing on the selection component of idea evaluation and to explore the occurrence of potential cognitive biases, such as an undue preference for adequacy over originality in creative ideas.

Alongside assessments of creative abilities and performance, 111 participants completed an idea production task, a subjective value rating task, a binary choice task and a second rating task, this time regarding the ideas' originality and adequacy level. 40 participants of our sample underwent this experiment in an MRI scanner.

At the behavioral level, we used computational modeling to explain individual ratings using the following utility function:

$$\text{subjective value}_{idea} = (\text{alpha} \times \text{Originality rating}_{idea}^{\text{delta}} + (1-\text{alpha}) \times \text{Adequacy rating}_{idea}^{\text{delta}})^{1/\text{delta}}$$

This model revealed that individuals followed different valuation patterns depending on the task: in their subjective value ratings, individuals integrated originality and adequacy in a balanced manner (alpha~0.5); but in their choices, individuals unreasonably preferred adequacy at the expense of originality (alpha<0.5).

Interestingly, the size of this bias varied among individuals, and those with higher biases against originality scored lower on creativity scores than those with lower biases (correlation between bias and creativity score: r=0.52, p=0.025), highlighting the interplay between valuation patterns and creative abilities.

These findings prompt consideration of which valuation patterns prevail in creative thinking: when producing ideas, do individuals follow their declarative, unbiased preferences (as in their ratings) or their implicit, biased preferences (as in their choices)? Current analyses including model comparisons of BVS signal during the idea production task are addressing this question.

Overall, this study challenges current accounts of the neurocognitive bases of decision-making during creativity and underscores the importance of employing neuroeconomics methods to explore creativity and its cognitive biases.

PS.02.03 - Psychological interventions increase motivation to exert effort to mitigate climate change

Jo Cutler [1], **Luis Contreras Huerta** [2], **Boryana Todorova** [3], **Jonas Nitschke** [3], **Katerina Michalaki** [4], **Lina Koppel** [5], **Theofilos Gkinopoulos** [6], **Todd Vogel** [1], **Claus Lamm** [3], **Daniel Västfjäll** [5], **Manos Tsakiris** [4], **Matthew Apps** [1], **Patricia Lockwood** [1]

[1] University of Birmingham, [2] University of Oxford, [3] University of Vienna, [4] Royal Holloway University of London, [5] Linköping University, [6] Jagiellonian University

Details

To limit the devastating effects of climate change, it is vital individuals engage in pro-environmental behaviours. Psychological interventions could be effective at promoting such actions. However, previous work has measured pro-environmental attitudes, missing a key aspect: effort. As pro-environmental actions are often more effortful, assessing the efficacy of interventions must include whether they increase willingness to exert effort for the environment. Here, we developed the novel Pro-Environmental Effort Task (PEET) and tested the effectiveness of 11 psychological interventions.

Methods
A large sample of participants (n=3,055), six representative samples from six countries across three continents, completed the PEET online. Participants chose to rest or 'work', rapidly clicking boxes on the screen, to earn rewards for a climate charity or a closely matched non-environmental cause. The effort required (50, 65, 80 or 95% of boxes clicked in calibration) and reward available (4, 12, or 20 credits) varied independently, enabling us to apply computational modelling to precisely quantify pro-environmental motivation.

Results
Strikingly, in the absence of an intervention participants were less motivated to exert effort to help the climate. Pro-environmental benefits were discounted by effort to a greater extent than donations to the non-environmental cause (OR=0.97, p=0.007). However, we found a small number of interventions increased motivation to benefit the climate, relative to the control charity. Specifically, interventions that targeted psychological distance and system justification significantly reduced the bias against helping the environment, across multiple measures and control analyses (ps<0.044). Computational modelling revealed these interventions affected devaluation of pro-environmental benefits by effort, independent of decision noise. Finally, while motivation to help both causes was associated with lower

apathy and subjective effort, motivation to help the climate was specifically associated with belief in climate change and pro-environmental policy support (*ps*<0.001).

Conclusion
Our results highlight the importance of tasks with real effort and incentives to precisely measure motivation to engage in effortful pro-environmental actions. Using a novel task, we reveal the individual predictors of pro-environmental motivation and identify psychological interventions that promote relative willingness to exert effort to protect the planet.

Poster Spotlights III

PS.03.01 - The anterior insula encodes salience in a subjective fashion and preferentially in the aversive domain

Jae-Chang Kim [1], **Stephan Nebe** [1], **Lydia Hellrung** [1], **Philippe Tobler** [1]

[1] University of Zürich

Details

The insula has been associated with processing salience (Molnar-Szakacs & Uddin, 2022) and salience prediction errors (Corlett et al., 2022). However, the exact form (e.g., objective vs. subjective) and extent (e.g., preferentially in aversive domain or domain-general) of salience processing by the insula remains to be determined.

To address these questions, we used a Pavlovian conditioning task (n=41; 22.4 ± 0.43 years, mean ± SEM; 19 women) involving appetitive, aversive, and neutral liquids. We behaviorally matched the absolute value of appetitive and aversive liquids by bidding (Becker-DeGroot-Marschak auction) and rating (general labelled magnitude scale), and used different cues to predict the liquids with different probabilities (p=0, 0.5, 1). In each trial, participants rated both the cue and the outcome, which allowed us to directly determine subjective salience prediction errors, the unsigned difference between the outcome rating and the cue rating within the MRI scanner. By contrast, objective salience prediction errors were defined by probabilities predicted by cues and actual outcomes.

A non-parametric analysis revealed a significant objective salience prediction error signal for valenced outcomes (i.e., appetitive and aversive) but not for non-valenced liquids (i.e., neutral) within the insula at the time of the outcome (p<0.05, FWE-whole brain voxel-level corrected). A parametric approach replicated these results but direct comparisons revealed a significantly stronger association of anterior insula activity with subjective than objective salience prediction errors at the time of the outcome. Moreover, subjective salience prediction errors activated the anterior insula also at the time of cues. As one would expect based on discounting, the cue-related signals were significantly weaker than the outcome-induced signals. Finally, separate analyses for the different domains (i.e., appetitive, aversive, neutral, and no outcome) revealed domain-general subjective salience prediction error signals in the anterior insula at the time of the outcome. By contrast, at the time of the cue, the coding of subjective salience prediction errors was limited to the aversive domain, compatible with weaker discounting of salience in the aversive than the appetitive or neutral domain.

Together, the anterior insula encodes an error in the prediction of subjective salience at each moment in time and discounts future salience to a lesser extent in the aversive than the appetitive domain. These findings reconcile two hitherto apparently conflicting views on insula function (salience coding common to appetitive and aversive domain vs. preferential coding of aversive events).

PS.03.02 - Bounded rationality in rats is accounted for by value normalization

Margarida Pexirra [1], **Jeffrey Erlich** [2]

[1] University College London, [2] NYU Shanghai

Details

Models of choice behavior typically rely on fundamental assumptions of rationality based on well-defined economic axioms. The Independence of Irrelevant Alternatives (IIA) axiom is one of these premises of rationality, stating that the relative preference between two options should be unaffected when extra options are added to the choice set. However, studies across different species have demonstrated violations of the IIA axiom, in a way that is consistent with the predictions of divisive normalization. To investigate the neural mechanisms that produce violations of the IIA axiom under a value normalization framework, we developed a task where rats make decisions between six options that differ in reward value. A trial was initiated when rats enter their snouts in a central nose port, triggering lights that indicate which of the six options were available for choice on a given trial. Forty-one different trial types were presented in an interleaved manner during a session: six where only one port was available, fifteen pairs, and twenty triplets. After a variable amount of time, a sound go-cue was played, releasing animals from fixation and allowing them to make a choice towards any of the illuminated available ports. Once selected one of the available ports, a volume was sampled from the reward distribution associated with the the chosen port and delivered via a dedicated reward port. This task structure, featuring the same option in multiple choice contexts, allowed us to test how the presence and value of a third (unchosen) option influenced choice ratios - a key metric of violations of the IIA axiom. Animals learned to choose the highest value port available on a trial-by-trial basis: developing stable and transitive preferences between two ports. Interestingly, in trials with a third irrelevant option (distractor), animals almost never chose the distractor, but the target choice ratio between the top two options was influenced in a way that was consistent with value normalization. Having established a framework for studying value normalization in rats, we are now collecting neurophysiological data from regions involved in economic decision-making, such as the orbitofrontal cortex and the frontal orienting field and planning causal perturbation studies to investigate how and where value normalization is implemented.

Poster Presentations

Poster Session I

P1-A-1 - Are consumer financial decisions too hard? The case of credit cards.

Michelle Lee [1], **Carsten Murawski** [1], **Nitin Yadav** [1]

[1] University of Melbourne

Details

Objective: As of 2023, the United States has accumulated a staggering $1.13 trillion in credit card debt, with delinquency rates on the rise, particularly among younger credit cardholders (Federal Reserve Bank of New York, 2024). Many credit cardholders could reduce their borrowing costs by choosing a more suitable credit card. Yet, selecting an appropriate credit card is challenging, even when the number of options to choose from is limited and future spending and repayment patterns are certain. Here, we characterise the computational complexity, hereafter simply complexity, of the credit card choice problem and test whether the complexity of individual credit card choices affects choice quality. **Method**: 26 participants (mean age = 24.42) completed a two-alternative forced choice task. Given a fixed 12-month spending and repayment schedule, participants needed to select the credit card that minimised their total borrowing costs. Each credit card had up to two cost features: an annual fee and an annual interest rate, which were sampled from the Australian credit cards market. In this study, the complexity of choosing a credit card was operationalised in two ways: the number of compounding steps required to compute the total interest payments and credit card feature composition. Complexity was higher with more compounding periods and when a credit card had an interest rate. We also measured subjective complexity based on participants' perception of task difficulty. **Results**: We found that as the complexity of a particular credit card choice increased, participants spent more time (β=0.39, $p<0.001$) and made more mistakes (odds ratio (OR)=0.12, $p<0.001$), but only when the cost of mistake was less than AU$114 per annum ($p<0.001$). In other words, complexity mattered, but not when one of the options was clearly better. Subjective complexity was positively correlated with more complex feature composition, but, surprisingly, it was inversely related to the number of compounding steps (OR=0.52, $p<0.001$). This suggests that participants relied on heuristics, leading to misestimation of the complexity of the choice problem. **Conclusion**: We introduce a framework to quantify the complexity of consumer financial decisions based on the computational resources required to make the decision. Using parameters from existing credit cards in Australia, we show that higher complexity choices can impair choice quality and incur non-trivial costs. Given that credit cards are considered to be one of the simplest financial products, our results suggest that many important financial decisions are beyond consumers' cognitive capacity.

P1-A-2 - Predicting entrepreneurial pitch success using behavioral, textual, and EEG neuroforecasting

Jin Ho Yun [1], **Sohvi Heaton** [2], **Peter Klein** [2], **Michael Platt** [1]

[1] University of Pennsylvania, [2] Baylor University

Details

Entrepreneurs often face difficulty pitching novel ideas to potential stakeholders, including suppliers of financial capital. How do they engage potential funders under conditions of uncertainty? Existing research in entrepreneurial finance relies mainly on self-reported subjective assessments which may not fully capture how a pitch shapes investor decisions in real time. We go beyond this literature by applying a neuroforecasting approach that integrates self-reports, textual sentiment analysis, and neural

responses via EEG to forecast the outcomes of entrepreneurs' pitches. We conducted a mixed-methods analysis of investor reactions among 28 subjects who watched 14 Shark Tank pitches. First, we analyzed EEG frontal alpha asymmetry (FAA), a biometric measure of positive motivation (from power spectral density) and neural synchrony or intersubject correlation (a measure of shared attention, emotion, and mindset) from time-frequency representations. Second, we employed natural language processing to analyze pitch transcripts for various sentiments. After each pitch video, participants self-reported their interest in obtaining further information, willingness to invest, perception of the pitch's quality, and expected financial returns. We also measured individuals' tendencies toward dynamic capability (sensing and seizing investment opportunities). We then used hierarchical linear models to predict investor interest with predictors including self-reported pitching content quality, investors' dynamic capabilities, and FAA (averaged whole span of each pitch). We found that all measures significantly predicted investor interest. Importantly, incorporating FAA and dynamic capability measures into the model significantly improved predictions, even controlling for age and gender. However, incorporating sentiment analysis from textual data did not increase variance. Second, the FAA predicted out-of-sample investors' interest (N=497) within the first 10 seconds of pitching, providing a biomarker of early impression and interest. Moreover, peak neural synchrony around 50 seconds into the pitch, when entrepreneurs were typically describing their product solutions, significantly predicted real-world deal outcomes. Thus, integrating EEG data with self-reported measures provides a robust predictive approach to evaluating and refining pitches to maximize successful outcomes. Our findings build upon recent work in neuroforecasting predicting pitch success using portable EEG, thus increasing our understanding of investor reactions to novel ideas and offering a practical solution for entrepreneurs to refine their presentation strategies and successfully secure a deal.

P1-B-3 - Neuroforecasting social media engagement with wildlife imagery

Tara Srirangarajan [1], **Cynthia Wu** [1], **Nik Sawe** [1], **Tierney Thys** [2], **Brian Knutson** [1]

[1] Stanford University, [2] California Academy of Sciences, National Geographic

Details

Visual imagery can motivate support for wildlife species, but the mechanisms underlying their impact remain unclear. To understand how images of wildlife species elicit engagement, we investigated links between behavioral probes, neuroimaging data (n=37), and social media engagement data (from National Geographic's Instagram content stream, one of the largest and most influential, with nearly 300 million followers as of 2024). As hypothesized, activity in brain circuits associated with anticipatory affect and value integration (i.e., the Nucleus Accumbens or NAcc, and Medial PreFrontal Cortex or MPFC) predicted individuals' engagement with and donations to depicted species (56 image stimuli). Group brain activity in the value integrative component of this circuit (MPFC) also forecast aggregate engagement on social media (i.e., total number of likes divided by number of current followers on Instagram). Further, MPFC activity correlated with activity in distinct circuits implicated face processing (i.e., the Fusiform Face Area or FFA) and mentalizing (i.e., the TemporoParietal Junction or TPJ). Finally, a neurally-inspired model derived from these correlated components forecast image engagement out-of-sample (276 additional image stimuli). These findings not only practically extend neuroforecasting to online engagement with wildlife images, but also theoretically imply that neural data can support the

discovery of stimulus features which promote online engagement. These findings therefore have applied implications for accentuating or altering image features to promote support for wildlife.

P1-B-4 - Untangling self-report and brain measurements in aggregate preference prediction

Hang-Yee Chan [1], **Christin Scholz** [2], **Emily Falk** [3]

[1] King's College London, [2] University of Amsterdam, [3] University of Pennsylvania

Details

A spate of studies in communication and consumer neuroscience have shown that neural activity from small groups of individuals tracked out-of-sample preference in aggregate (i.e., neuroforecasting), akin to extrapolating opinion polls or customer surveys to predict market outcomes. What do aggregate brain signals reveal where self-reports do not?

In a mega-analysis (i.e., pooled analysis of raw neuroimaging data) of 16 functional MRI studies involving images, texts and videos from various domains such as entertainment, marketing and health communications (572 participants, 768 stimuli, 23,085 trials), we studied how aggregate neural activity during stimuli exposure tracked out-of-sample outcomes associated with the stimuli (clickthrough rates, liking, etc), and whether such brain-to-market effect was mediated via aggregate self-report preference of the neuroimaging sample.

In brief, (a) neural activity implicated in reward and mentalizing preceded self-report preference within individuals, and tracked out-of-sample aggregate preference; and (b) the predictive effect of reward-related (but not mentalizing) activity of the neuroimaging sample on out-of-sample preference is mediated by within-sample self-reports.

Our findings suggest that neuroforecasting works by reflecting the self-report preference of the sample, while at the same time it potentially adds value by extracting the social relevance of stimuli not captured by existing self-report measurements.

P1-B-5 - NeuroChic: Unraveling Sustainable Fashion Advertising through Brainwave Insights

Elena Gasulla Tortajada [1], **Paulo Duarte** [2], **Susana C. Silva** [3], **Jose Paulo Marques Dos Santos** [4]

[1] Universidade Beira Interior, [2] Universidade da Beira Interior, [3] Universidade Católica Portuguesa, [4] Universidade da Maia

Details

Despite extensive research on the effect of ad appeals on sustainable fashion consumption (Dhir et al., 2021; Grazzini et al., 2021; Legere & Kang, 2020; Rausch & Kopplin, 2021), reliance on explicit self-report measures limits understanding of consumers' implicit mental processes, leaving a gap in linking attitudes

to real behavior (Hassan et al., 2016; Rausch & Kopplin, 2021). This study advances prior research by studying consumer brain responses to different types of sustainable fashion advertising content and how arousal can predict actual purchasing behavior.

An electroencephalography (EEG) experiment with 40 participants will analyze the effects of sustainable fashion advertising appeals on consumer cognition, examining the distinctions between emotional and informative content, in line with Guitart & Stremersch (2021) definition. Self-reported measures of purchase intention towards the advertised products will be collected before and after exposure to the stimuli, and the responses will be compared with the EEG results to identify disparities between stated attitude and neural activity. Additionally, the correlation between the neural responses and actual purchasing behavior, measured by click-through rates in connection with the ads, will be investigated in a large population of 500 respondents in a consumer panel.

Our findings should reveal that emotional and informative ads activate distinct brain regions, correlating with varied effectiveness in influencing consumer behavior. Emotional stimuli are likely to engage areas tied to affective processing, such as frontal alpha asymmetry. At the same time, informative content is expected to stimulate regions associated with cognitive processing and willingness to pay, like theta waves and the beta over alpha plus theta ratio, along with differences in gamma powers. Participants exposed to emotional ads are anticipated to self-report greater purchase intention toward sustainable fashion products compared to those exposed to descriptive ads. However, on a broader scale, ads featuring informative content are expected to drive higher real consumer behavior.

The study´s results are expected to demonstrate EEG measures' superior predictive power over overstated preferences in understanding consumer behavior toward sustainable fashion campaigns. The results will contribute to emphasize the crucial role of tailored advertising to bridge the gap between attitude and behavior in sustainable fashion consumption. These insights can guide marketers and policymakers in crafting more impactful communication strategies and policies aligned with consumer values, promoting sustainable fashion adoption.

P1-B-6 - Unraveling other's preferences in bargaining through eye movements

Mrugsen Gopnarayan [1], **Sebastian Gluth** [1]

[1] University of Hamburg

Details

Eye movements are intricately coupled with decision-making, serving as a window into the attentional processes of decision-makers. In social settings, we intuitively use this principle to infer others' preferences while they are deciding. Previous work has demonstrated that showing the gaze pattern of one participant to another in a coordination game enables them to better understand each other's preferred choices. In the present study, we use this capability to test whether a seller can reach better agreements with a buyer by seeing their eye movements in a cooperative bargaining setting. The task involves the seller attempting to sell 20 different products with multiple attributes (e.g., cars, laptops, washing machines) to a buyer. The seller's goal is to determine the relevance of different attributes to the buyer in order to offer (and ultimately sell) the appropriate products. For example, when choosing

among cars with attributes speed, safety, and carbon emission, a climate-conscious buyer would primarily be concerned with the carbon emissions of the car, whereas another buyer might prioritize speed. In our task, the seller must deduce such preferences from different sources of information: the buyer's choices, response times, and eye movements.

To test whether sellers can use buyers' eye movements to infer their preferences, we will divide the sellers into two groups: a Control group (N=36 pairs), which will have access only to reaction times and choices, and an Experimental group (N=36 pairs), which will also have access to the buyers' eye movements. Our hypothesis is that sellers in the Experimental group will be able to make better predictions about buyer preferences and achieve more and quicker sales. Additionally, we aim to test whether response times can be used to infer buyer preferences.

Based on the analysis of pilot data for the Control group (N = 10 pairs), the buyers' response times are predictive of their prespecified preferences ($R = 0.26$, $p < 0.05$). The seller's updates in consecutive offers are weakly correlated with the buyer's response times ($R = 0.061$, $p < 0.05$). We plan to complete data collection before the conference to provide a comprehensive analysis of how eye movement data can optimize bargaining strategies and enhance decision-making processes in social interactions.

P1-B-8 - Health goals alter the neural encoding of specific food nutrients

Rémi Janet [1], **Matthew Bachman** [2], **Cendri Hutcherson** [2], **Anita Tusche** [1]

[1] Queen's University, [2] University of Toronto

Details

BACKGROUND: Dietary choices are, at least in part, computed by considering the nutritional properties of a food item (e.g., protein, fat, carbohydrates, sodium, and vitamin content). Yet, surprisingly little is known about how the brain encodes food's nutritional profiles and how malleable these representations are. **OBJECTIVES**: Which brain areas contribute to the construction of foods' nutritional profile during dietary choice? How do regulatory goals alter the neural representations of foods' nutritional components? **METHODS:** To address these questions, we used data from two previously published fMRI studies (N=95). All participants performed an established laboratory food task under different experimental conditions (including food choices in natural and health-focused, regulatory settings). Using multivariate analyses of functional MRI data, we identified several brain areas that allowed decoding the elemental nutritional components of foods while participants made their dietary choices (i.e., protein, fat, carbohydrates, sodium, and vitamin content). Brain areas predicting objective nutrient values included, among others, the dlPFC, vlPFC, lOFC and mOFC. Second, we found goal-consistent changes in the neural representation of nutritional values when participants focused on health goals (e.g., in the dlPFC). Third, participants' weight status was linked to differences in neural representations of nutritional components during dietary choices. **CONCLUSIONS**: By showing how nutritive information is represented in the brain, our findings provide a mechanistic account for the construction of food value guiding dietary choices. Moreover, our results shed light on how the brain copes with the challenge of varying dietary goals, capitalizing on goal-consistent changes of neural representations of elemental nutritional features. Given the potential role of food valuation and nutritional information in the

development of dysregulated eating in obesity or anorexia, these findings offer novel insights into potential avenues for promoting healthier eating behaviors.

P1-B-9 - The effect of commercials on the neural signal of value

Inbal Gur Arie [1], **Dino Levy** [1], **Adam Hakim** [1]

[1] Tel Aviv University

Details

A basic aim of marketing research is to predict consumers' preferences and the success of marketing campaigns at the population level. Traditional marketing tools, however, have limitations, necessitating novel measures to improve predictive power. In a previous study, we developed a deep learning model that predicted subject-specific preferences of pictures of consumer products based on their EEG data. In the current study, we extend this method to video commercials.

There were three parts to the current experiment. In the first part, subjects (n=113) viewed a series of different consumer products and indicated their willingness to pay (WTP) for each product using a classic BDM task. Thereafter, in the second part, subjects viewed commercials for these same products, and indicated, again, their WTP for each of the products, and ranked how much they liked the commercial itself. During these parts, we recorded subjects' brain activity using EEG. A week later, in the third part of the experiment, subjects again indicated their willingness to pay for the products, but without EEG recordings, outside the lab, as an online experiment. We found that watching commercials increased subjects' WTP, but only immediately after watching the commercial of the product. After a week, subjects' WTP returned to baseline. Furthermore, commercials' liking and products' WTP are only weakly correlated ($R^2 = 0.098$, $p < 0.001$), indicating they represent different aspects of commercial preference. Moreover, we found that traditional EEG frequency bands are significantly, but weakly, correlated with subjects' WTP measured while watching the products' pictures ($R^2 = 0.017$, $p < 0.001$) or while watching their commercials ($R^2 = 0.027$, $p < 0.001$). When we conducted a logistic regression using the EEG frequency bands to predict subjects' WTP we could not predict successfully over chance levels. However, using our deep learning network, we were able to predict subjects' preferences while watching products' pictures with an average accuracy of 0.73 on out-of-sample subjects. Furthermore, the network predicted subjects' WTP while watching a commercial at an averaged accuracy of 0.73 and 0.66 on out-of-sample subjects. The network was even able to predict subjects' preferences a week later, at an averaged accuracy of 0.66.

To conclude, people's preferences seem to generally be influenced by relevant stimuli as commercials, while their liking and product value do not necessarily go hand in hand. Additionally, while traditional EEG features did not give rise to the full picture of value's neural correlates, using advanced non-linear methods, provided far better prediction accuracies

P1-C-10 - An investigation of learning in ambiguous contexts

Sibilla Di Guida [1], **Luca Polonio** [2], **Davide Marchiori** [3]

[1] IMT School for Advanced Studies Lucca, [2] University of Trento, [3] IMT school for advanced studies Lucca

Details

Study's objectives

It is commonly accepted that full, immediate, and accurate feedback in repeated strategic interactions should correct peoples' biases and beliefs in favor of the rational course of action. With this study we challenge and qualify this view. We test whether in ambiguous albeit accurate information settings, people bend the interpretation of the feedback received to reinforce some preexisting heuristic of play, so that learning outcomes that are apparently compatible with optimal behavior mask instead non-rational strategic intents

Methods used

We study this effect in an experiment in which participants faced sequences of different 2-person 3x3 matrix-form games designed to embed five of the most observed and studied heuristics of play: level-1, level-2, focal point, optimistic, best reply to optimistic. We defined a pair of treatments for each target heuristic: one including sequences of games in which the equilibrium and heuristic action overlap (labeled "ambiguous"), and one sequences of games in which these actions are distinct ("unambiguous"). In total, we tested 10 treatments. Data were collected online on the platform "Prolific", with 826 subjects from the UK sample, randomly allocated across treatments

Results obtained

Results show in detail that when learning occurs in settings in which equilibrium and target heuristic actions overlap, learning equilibrium play is substantially more difficult, and people rather use the ambiguous and immediate feedback information to reinforce behaviors that are compatible with the target heuristic. Additionally, we show that this "distraction from equilibrium" effect varies substantially across the different target heuristics considered

Conclusions

Our work contributes to the understanding of the interplay between feedback information and peoples' internal interpretation of a strategic situation. The results have two kinds of implications: theoretical and methodological. From a theoretical perspective, we show that immediate and accurate feedback is not sufficient for correcting beliefs and controlling internal interpretations of the game situation, and that such info, although fully transparent, can instead be used to reinforce behaviors other than rational. From a methodological perspective, our results suggest caution in the interpretation and generalization of data from learning experiments. They specifically caution against mistakenly equating an increase in the frequency of equilibrium choices (what we observe) with what individuals actually learn (which we cannot observe, but can still infer, although only through purposely designed experiments), a common practice in economics research

P1-C-11 - Examining cognitive and motivational factors influencing optimal decision-making through eye-tracking

Mattia Adamo [1], **Sohvi Heaton** [2], **Giorgio Coricelli** [3]

[1] IMT School for Advanced Studies Lucca, [2] Baylor University, [3] University of Southern California

Details

OBJECTIVE: Variations in managerial decision-making frequently result in diverse outcomes, impacting the heterogeneity in firm performance. Hence, it is essential to comprehend the conditions under which optimal decisions are made, particularly within the context of strategic interactions. We analyze data from the eye-tracking experiment of Polonio & Coricelli (2019). While the previous study observed the decision making patterns in a game theoretical perspective aimed at categorizing all types of decision makers in the behavioral economics realm, we extend the results under a managerial decision making lens whose objective is to explain the differential performance in terms of best responding to beliefs on the opponent's action as the difference between individuals endowed with managerial dynamic capabilities and not. We draw results from the experiment by taking into account factors such as the sophistication level of the decision model (Nagel, 1995), lookup patterns, and prosocial motivation. **METHODS**: Participants (n=72) completed a choice task and a belief elicitation task. During the choice task, participants engaged in 18 3x3 strategic games that could be solved in varying steps (2, 3, or 4) of iterated elimination of dominated strategies. Participants assigned probabilities to the possible actions of their opponents. Choices and responses were categorized based on level-1, level-2 reasoning, and Prosociality models. The top performers were identified as those who more consistently aligned their decisions with their beliefs rather than the strategies of other players. Eye-tracking methodology was used to track players' gaze trajectories on their own and others' payoffs. **RESULTS:** We found that level 2 players who best respond to the belief that their counterpart is choosing the Level-1 action generally align their decisions with their beliefs in one-shot games ($p < 0.01$). Their performance is particularly strong when combined with a moderate number of fixations on others' payoffs and a preference for prosocial behavior ($p < 0.01$). Considering prosocial choices alone is a significant negative predictor of best responses ($p < 0.01$). **CONCLUSIONS:** Our analysis reinforces the notion that individuals vary in their levels of cognitive sophistication, social preferences, and attention allocation during the decision-making process, leading to heterogeneity in strategic interactions. In a broader sense, our study contributes to the game theory literature by offering empirical evidence on the Level-k model. Furthermore, our research offers practical implications for managers tasked with making strategic decisions for their organizations.

P1-C-12 - Trust, Risk, and Social Comparison

Shulang Yue [1], **Iris Vilares** [1]

[1] University of Minnesota, Twin Cities

Details

Numerous fields have demonstrated that our decisions are guided by social information, and one important piece of social information is social comparison. However, virtually all prior studies investigating the influence of social comparison on trust decisions employed priming and performance based social comparison as comparison direction paradigms. These paradigms are often concerned with personal meanings including one's abilities, skills, or achievements. Nevertheless, we frequently find ourselves in specific social positions for no obvious reasons. Building on the work of Moyal et al. (2020), the current study aims to investigate how trust, trustworthiness, and risk-related decisions are made subsequently in unrelated situations following exposure to arbitrary-based social comparison contexts. In addition, the study aims to directly compare people's behaviors in trust and risk environments under different social positions in a single experimental paradigm. The study employed a 3 (Comparison direction: upward, lateral, downward; between subject) by 2 (Economic game: trust, risk; within subject) mixed design. A power analysis determined that a sample size with at least 252 participants is needed to detect a medium-sized effect ($f = 0.25$) with a 95% power. Participants completed three tasks: a social comparison induction task, a one-shot Trust Game, and a one-shot Risk Game. The social comparison induction task was always completed first, and the order of the Trust and Risk games was counterbalanced. Our results (N = 302) showed a significant difference of how decisions were made under trust and risk environments ($p < 0.001$). In general, participants transferred smaller amounts in the Risk Game than in the Trust Game, and we observed a pattern in which making downward social comparisons resulted in less subsequent risk-taking behavior, while making upward social comparisons resulted in an increased tendency to take risks. However, while no significant difference was observed for the effect of arbitrary-based social comparison in trust, trustworthiness, and risky decision making, a main effect of social comparison was observed in allocation decisions ($p < 0.001$) and perceived relative rank ($p = 0.003$). Overall, our findings suggest that people's subsequent trust, trustworthiness, and risky decision making appear to be less affected by exposure to randomly manipulated social comparison contexts than expected, but they perceive risk and trust environments differently.

P1-C-13 - Mentalising underlies strategic coordination in Guinea baboons (Papio papio).

Toan Nong [1], **Nicolas Claidière** [2], **Edmund Derrington** [1], **Joel Fagot** [2], **Jean-Claude Dreher** [2]

[1] Université Claude Bernard Lyon 1, [2] National Center for Scientific Research

Details

It remains controversial whether the ability to mentalise is confined to humans. To address this question, Guinea baboons living in a social colony freely came to play a two-players coordination game with any other baboon, or alone (social vs solo conditions). Although in both conditions they interacted with an identical Artificial Agent, they achieved better coordination in the social than in the solo condition. A mentalising computational model integrating relative dominance ranks accounted for baboons' coordination behaviour better than computational models without mentalizing components. These findings characterize the computational mechanisms used by the baboon's brain for coordination learning and demonstrate that a basic form of mentalising may have evolved in the common ancestor of the Old-World monkey and apes that diverged 30 million years ago.

P1-D-14 - Perceptual underpinnings of individual risk attitudes in adolescent dyscalculics

Maike Renkert [1], **Gilles De Hollander** [1], **Karin Kucian** [2], **Christian Ruff** [1, 3]

[1] University of Zürich, [2] University Children's Hospital Zurich, [3] University of Zurich

Details

Recent theories of decisions under uncertainty suggest that (parts of) the variability in individual risk attitude can emerge from differences in the perception of numerical magnitudes: Risk aversion might partly reflect a perceptual bias leading to underestimation of the larger magnitudes of risky payoffs [Khaw, 2021]. This novel theoretical account is informed by extensive literature on the so called 'number-sense' - an innate cognitive system housed in parietal cortex that helps us perceive and represent numerosities. The acuity of this neurocognitive system shows large individual variability and has been linked to variability in apparent risk preferences [Baretto-García, 2023]. In the current study we directly test this hypothesis in adolescents with Dyscalculia, a number-and math-specific developmental learning disorder that has been suggested to originate from an imprecise number sense.

We recruited 33 adolescents diagnosed with dyscalculia and 33 neurotypical controls matched for age and gender (15-23 years, 28 females per group). Participants underwent a comprehensive behavioral and neural assay comprising anatomical MRI scans, fMRI scans during a magnitude comparison task, a risky choice task, a set of cognitive tests, and multiple questionnaires. We analyzed the behavior from the magnitude task with a Bayesian perceptual choice model to disentangle the underlying cognitive components.

Model fits revealed that dyscalculics were substantially worse than control subjects at keeping numerical information in working memory ($p=0.004$), while at the same time focussing more on directly accessible information on the screen when making the decision ($p=0.019$). This working-memory impairment aligned with lower scores in independent tests of visuo-spatial reasoning and spatial short-term memory ($p<0.016$), validating the model's assumption. In the risky choice task, a measure of risk attitudes (indifference point; payoff ratio at which subjects appear indifferent between risky and safe option) strongly depended on stake size (more risk seeking for larger stakes). This effect was significantly more pronounced in dyscalculics ($p=0.022$), showing that dyscalculics take magnitudes differently into account when making risky decisions. In upcoming work, we will link these behavioral findings to neural data, specifically to how numbers are represented in the parietal magnitude processing system and how this information is integrated into whole-brain fronto-parietal networks to inform decision making. With this, we hope to improve our understanding of neural processing peculiarities in dyscalculia and their everyday-life implications for economic decisions.

P1-D-15 - A gut feeling: Exploring the Effects of Probiotics on Risk-Taking Behavior using TMS and EEG

Aline Dantas [1], **Elisabeth Bruggen** [2], **Peiran Jiao** [2], **Alexander Sack** [2], **Teresa Schuhmann** [2]

[1] Maastricht University & Donders Institute, [2] Maastricht University

Details

The gut-brain axis (GBA) involves communication between gut bacteria, the enteric nervous system, and the central nervous system, impacting brain activity and behavior. Limited research exists on how this axis affects human decision-making and its underlying neural processing. Furthermore, the relevance of potential key areas, such as the ventromedial prefrontal cortex (VMPFC), in the GBA interaction during decision-making remains largely unexplored.

This study's **objective** is to further explore the influence of the gut-brain axis (GBA) on the risk-taking behavior of healthy adults with the use of probiotics/placebo, transcranial magnetic stimulation (TMS), electroencephalography (EEG) and the Maastricht Gambling Task (MGT).

The **methodology** used includes double blinded mixed design, combined with the prolonged intake of probiotics/placebo (30 days). The study includes two sessions, before and after the intake of probiotics/placebo, during which participants underwent three inhibitory TMS protocols (VMPFC, superior parietal lobule (SPL) as control location and sham) in randomized order.

The **results** obtained show significant increases in risky choices ($t(44360) = 2.02$, $p = 0.043$) and response times ($t(43372) = 5.28$, $p < .001$) after prolonged intake of probiotics, compared to placebo. EEG analyses show significant increases in both right (beta = 0.08, $t(619) = 3.13$) and left alpha-power ($t(628) = 2.86$, $p = 0.004$) for females only and significant reductions in both right (beta = -0.12, $t(619) = -2.90$, $p = 0.004$) and left alpha-power ($t(628) = -3.09$, $p = 0.002$) for males after the probiotics protocol. VMPFC stimulation led to significant increases in risk-taking behavior ($t(44351) = 2.09$, $p = 0.036$). No other significant effects were observed.

In **conclusion**,these findings confirm the influence of the GBA on risk-taking behavior, with significant increases in risky choices after prolonged probiotics intake. Such effects cannot be attributed to more automatic responses, considering the significant increases in response times after probiotics intake. Nonsignificant interactions between probiotics and VMPFC deactivation indicate that this area does not play a fundamental role in the GBA interaction during risky decisions. Probiotics led to different EEG results for males and females, reinforcing recent findings that GBA manipulations might have gender-specific effects. This study adds important contributions to the investigation of the GBA's influence in human decision-making, confirming its influence on risk-taking behavior and adding important insights on its underlying neural processes.

P1-D-16 - Translational characterization of neural circuits underlying risky choice

Leili Mortazavi [1], **Felicity Gore** [1], **Karl Deisseroth** [1], **Brian Knutson** [1]

[1] Stanford University

Details

Despite the centrality of risky choice to behavioral theory and financial practice, neural mechanisms of risky choice have remained elusive. Human neuroimaging studies have reliably found that anticipatory

activity in distinct circuits predicts risk-seeking (Nucleus Accumbens; NAcc) and risk-avoidant (Anterior Insula; AIns) choices. Diffusion Weighted Imaging (DWI) data has further linked risk preferences to the coherence of a white-matter tract connecting the AIns and NAcc (Leong et al., 2016). Comparative structural studies have also identified glutamatergic projections from the AIns to the NAcc, implying a direct and rapid route of influence. Whether activity in this pathway can causally regulate risky choice, however, remains to be demonstrated. To gain mechanistic insight into the structural and functional properties of this tract, we probed AIns-NAcc projections in rats by combining DWI with viral tracing and projection-based optogenetics.

To maintain translatability to humans, we sought to use DWI as a parallel method to identify the AIns-NAcc tract in rats, and then to validate the tract with brain clearing methods. We scanned 15 rats with DWI and performed probabilistic tractography with whole insula as seed and a combined mask of NAcc core and shell as target. We consistently observed bilateral tracts connecting the anterior part of the insula to the lateral core of NAcc. We then mapped the whole-brain projection pattern of the AIns by injecting a fluorophore (AAV8-hSyn-oScarlet) into the AIns and performing whole-brain tissue clearing and light sheet microscopy. After registering the cleared data to diffusion tractography results, we confirmed the trajectory of the AIns-NAcc tract detected with DWI. We further found that performing whole-brain tractography with a sphere seed mask placed on the injection site yielded high convergence with whole-brain tract tracing results, particularly in short-range projections including the AIns-NAcc pathway. Finally, we injected an excitatory opsin (AAV8-hSyn-ChRmine-oScarlet) into the AIns and placed an optic fiber at the identified coordinates of its terminals in the lateral NAcc core to optogenetically probe the causal role of this projection on approach versus avoidance behavior. Pilot results (n=4) yielded preliminary evidence that stimulation of the AIns-NAcc pathway may be aversive and can inhibit approach behavior. Overall, these findings provide a methodological roadmap for translational research on neural mechanisms underlying risky choice, and highlight targets for causally manipulating risky choice in animals and humans.

P1-D-18 - Sugar liking as an indicator of reward sensitivity in delay discounting and risk-taking tasks

Anna Davidovich [1], **Ksenia Panidi** [1]

[1] National Research University Higher School of Economics

Details

Objective: Both risky and intertemporal choices involve reward valuation. Monetary choices people make when risk or time delay is present might be linked to individual reward sensitivity. Previous studies indicated that people with high preference for rewarding stimuli (e.g., sucrose) also demonstrate increased impulsivity, i.e. tendency to choose smaller immediate over larger delayed rewards. However, as time delays are inherently associated with increased risks this link might be potentially modulated by individual risk preferences, with higher risk aversion leading to preference for immediate over delayed reward. In this study, we use the sucrose solution liking test as an indicator of reward sensitivity to test two hypotheses – first, that reward sensitivity is linked to risk preferences, and second, that this link might modulate the association between reward sensitivity and impulsivity.

Methods: Participants (N=38) completed a set of behavioral tasks in a randomized order. Risk preferences were assessed with the lottery choice task involving binary choices between obtaining a certain amount of money versus playing a lottery with varying amounts and probabilities. The delay discounting task was used to evaluate reward impulsivity with binary choices between an immediate smaller and a delayed larger amount. To assess reward sensitivity the Kampov-Polevoy sweet-liking test was used, which involved tasting the sucrose solutions of 5 different concentrations and indicating subjective liking for each solution. Based on the sweet-liking test responses participants were divided into the sweet-liking and sweet-disliking groups. The link between sweet-(dis)liking and risky and intertemporal preferences was then tested via the regression analysis.

Results: Regression analysis revealed that sweet-likers were more sensitive to the time delay compared to sweet-dislikers – increasing the delay by one day increased the probability of choosing an immediate option significantly more for the sweet-likers group compared to sweet-dislikers. Further analysis revealed that this effect was explained by differences in risk preferences. Sweet-likers were significantly more risk averse compared to sweet-dislikers. Accounting for participants' risk preferences in the analysis of delay discounting choices eliminated the effect of sweet-liking on impulsivity.

Conclusions: The study provides novel evidence that higher preference for rewarding stimuli is linked to increased risk aversion, and that this link might explain previously found association between reward sensitivity and impulsivity.

P1-D-19 - Metacognition generalizes across perception, delay discounting, and risk and ambiguity preference tasks

Corey Plate [1], **Dhruv Govil** [1], **Mandy Renfro** [1], **Ricardo Pizarro** [1], **Charles Zheng** [1], **Zoe Boundy-Singer** [2], **Corey Ziemba** [3], **Silvia Lopez-Guzman** [4]

[1] National Institute of Mental Health, [2] University of Texas, [3] University of Texas at Austin, [4] National Institutes of Health

Details

Objectives: Metacognition is the process of reflecting on one's thoughts and behaviors. Metacognitive ability is often measured by modeling the relationship between confidence reports and behavior in tasks where performance can be objectively assessed. Previous work explored if this ability is conserved across tasks and different domains such as perception and memory. However, standard models of confidence are inappropriate for preference-based tasks where choices are not objective. As a result, it is unclear if metacognitive ability in perceptual decision-making extends to value-based decision-making. We apply a novel computational model of confidence that provides separable parameters for confidence bias and metacognitive ability, i.e., the uncertainty of one's own decision uncertainty — meta-uncertainty. We compare parameters across 3 tasks spanning perception and value.

Methods: 149 CloudResearch participants completed three online decision-making tasks: an orientation discrimination (OD) task (divided into two conditions differing in their contrast volatility), a delay discounting (DD) task, and a risk/ambiguity (RA) preference task, with trial-by-trial confidence reports.

The preference-based task behavior (DD and RA) was fit with a hyperbolic discounted utility model (Lopez-Guzman, et al., 2018) and a risk/ambiguity model (Gilboa & Schmeidler, 1989), respectively. Confidence reports from the DD and RA and from the two conditions of the OD were separately fit using a hierarchical Bayesian extension of the CASANDRE model (Boundy-Singer, et al., 2023) which generated estimates of metacognitive ability for each participant. We then tested the "within-domain" (perception-perception, value-value), and "across-domain" (perception-value) correlation in this ability.

Results: We found a strong correlation within the perceptual domain (ρ=0.6, p<0.001) when comparing meta-uncertainty from the high- versus low-volatility blocks. Similarly, our results show meta-uncertainty correlated across the value-based decision tasks (DD and RA, ρ=0.4, p<0.001). We observed a moderate correlation when comparing metacognitive ability across domains (OD-RA ρ=0.21, p<0.05, OD-DD ρ=0.24, p<0.05).

Conclusion: Our results suggest that metacognitive ability may be domain-general across value-based and perceptual decision-making. Our computational approach allows for the application of sophisticated measures of metacognitive ability even in preference-based tasks. This approach opens the door for use in psychiatry research where previous measures of confidence are promising, but limited by the fact that they are based only on perceptual decision-making.

P1-D-20 - Delay, probability and effort discounting: convergent and divergent measures of reward valuation

Suzanne Mitchell [1], **Deborah Sevigny-Resetco** [1], **Aaron Brah** [1]

[1] Oregon Health & Science University

Details

Objectives: The National Institute of Mental Health's Research Domain Criteria suggests that the subjective value of rewards can be examined using choice tasks in which reward alternatives differ in their size/quality and their delivery probability, delivery delay or required effort completion. However, it is unclear to what degree these three assessment techniques can provide converging information about subjective reward value for individuals or whether individual differences in the moderating effects of probability, delay and effort costs obscure such individual differences. Comparisons of the effects of these different costs on the devaluation of hypothetical monetary rewards was used to examine this issue.

Methods: Using a within subject design, healthy control participants were required to choose between smaller, hypothetical monetary rewards available for sure, immediately with no effort versus larger rewards available either probabilistically, or after a delay or following the successful completion of physical or cognitive effort. We collected summary measures of the extent to which preferences changed as reward probability declined, delay to reward lengthened and effort requirements increased (discounting indices). Response times were also collected, and, for some individuals, eye movements were recorded. A similar study assessing only preference and response times was completed with rats choosing between sucrose solution rewards.

Results: Regardless of species, correlations amongst discounting indices were low: individuals that devalued a reward because it was delayed did not necessarily devalue rewards that required substantial effort to obtain, etc. Further, response times and, for human participants, gaze characteristics, did not alter similarly as a function of changes in the degree of risk, delay, or effort. These results suggest that probability, delay, and effort costs moderate subjective value via somewhat independent mechanisms. Inclusion of multiple reward amounts ($10, $25, $50) allowed computational models to be created that include a term reflecting reward amount, but correlations between this term and discounting indices were small and differed for risk, delay, and effort, suggesting the subjective value of the reward amount was not independent of reward cost.

Conclusions: Interest has been growing in the neurobiology of anhedonia (pleasure), and ways to distinguish changes in behavior due to anhedonia from changes due to apathy (motivation). Assessing the subjective value of commodities independent of the costs associated with obtaining them is challenging using choice procedures alone, and thus may not provide a clear way to identify interventions that affect anhedonia and apathy.

P1-D-21 - Testosterone increases risk-taking when social status is at stake

Annabel Losecaat Vermeer [1], **Maarten Boksem** [2], **Christian Gausterer** [3], **Claus Lamm** [4]

[1] Leiden University, [2] Erasmus University Rotterdam, [3] Medical University of Vienna, [4] University of Vienna

Details

Objectives: The steroid hormone testosterone has long been associated with increase risk-taking, yet evidence supporting this association is mixed, suggesting other moderators or contextual factors may play an important role. A key function of testosterone is to promote status-seeking behavior, which could happen through risk-taking. An open question is under which circumstances testosterone causally affects risk-taking behavior, particularly in relation to its impact on social status challenges. To address this gap, we conducted a double-blind, placebo-controlled study utilizing both monetary gambles and social status outcome-based gambles. In addition, we explored potential moderators previously defined in testosterone research, such as baseline cortisol.

Methods: A total of 166 male participants took part in the study (testosterone: n = 82, mean age = 24.38 years, *SD* = 3.33), placebo: n = 84, mean age = 24.55 years, *SD* = 3.64). Participants performed a competition task in which they were assigned either a high or low status rank position. Next, they played two risk-taking tasks in fixed order: i) a risk task with gambles containing only monetary gain-loss outcomes without direct impact on their social rank, and ii) a second risk task with gambles directly affecting their social rank without monetary incentives.

Results: First, for monetary gain-loss gambles, we observed no effect of testosterone administration (vs. placebo) in risk-taking (drug treatment: $z = -0.52$, $p = .602$) nor was there an interaction with social rank (low vs. high) (drug x rank, $z = 0.84$, $p = .399$). However, participants of low (vs. high) social rank increased risk-taking, irrespective of drug treatment (rank: $z = 2.70$, $p = .007$). Subsequently, when participants faced gambles directly influencing their social status position, testosterone administration (vs. placebo)

enhanced risk-taking in those who had a high social rank (drug treatment x rank: $z = 2.64$, $p = .008$; for high social rank, OR = 3.85, $p = .036$; for low social rank, OR = 0.55, $p = .285$). Baseline cortisol did not moderate the relationship between testosterone and social rank in risk-taking (monetary gambles: $z = -0.50$, $p = .614$; status gambles: $z = -1.18$, $p = .237$).

Conclusions: Our findings suggest a context-dependent role of testosterone in modulating risk-taking concerning social status. Future research should directly compare risk-taking for both monetary and status outcomes using identical gambles within the same task. Our study provides novel insights for existing research in the field and theoretical models of testosterone in risk-taking as a manifestation for status-seeking behavior.

P1-D-22 - Greater value of resolving uncertainty and positive outcomes in self-relevant compared to social information seeking

Jo Cutler [1], **Matthew Apps** [1], **Patricia Lockwood** [1]

[1] University of Birmingham

Details

Objective

People make sense of the world by acquiring information and information about others has vital implications for interpersonal relationships and addressing global challenges. However, with ever-increasing information available, we must constantly choose whether to seek it. Two reasons to seek information are whether it resolves uncertainty and makes us feel good, depending on its valence. These benefits must be compared with costs, such as the time required. Existing work suggests people seek self-relevant information to resolve uncertainty, but the mechanisms of valuing social information are unknown. Here in a preregistered study, we tested whether the same factors drive social and self-relevant information seeking.

Methods

Two large online samples completed a novel information seeking task that measures choices to seek information about financial rewards for oneself or an anonymous other person (n1=238; n2=244, 229 tested again a week later). Information varied in uncertainty, valence, and the time required to seek it, but had no instrumental utility as participants could not affect outcomes. Participants also completed a discounting task that measures how they discount financial outcomes for self and other by the time to obtain them. This enables models that differentiate the values of outcomes and information. The task showed good test-retest reliability (ICC self=0.86, other=0.79) and all results replicated across samples (reported as n1 | n2 timepoint 1).

Results

Participants overall did wait to find out about outcomes for others but were significantly less interested in this social information than information about rewards for themselves (mean other=35% | 35%; self=48% | 45%; odds ratio (OR)=0.58 | 0.61, *p*s<0.001). While both higher uncertainty and more positive valence promoted information seeking for both recipients, these factors had a greater effect when outcomes affected the participant (uncertainty OR=0.94 | 0.95, valence OR=0.88 | 0.93, *p*s<0.015). In other words, participants valued social information less than self-relevant information because they were less interested in resolving uncertainty and seeing positive outcomes for others. Finally, participants were less likely to seek information about both self and other as the time cost increased (OR=0.84 | 0.79, *p*s<0.001).

Conclusions

Our results show that people value information about other people, and are willing to wait to seek it, depending on the same factors that promote seeking information about rewards for oneself. However, lower curiosity to resolve uncertainty about social compared to self-relevant outcomes leads to lower interest. The reliability of our novel task also suggests stability in individuals' preferences for information over time.

P1-D-23 - Experimentally-induced Anxiety Reduces Behavioral Dishonesty in Low-Trait Anxious Individuals: Neural Evidence from an EEG Study

Praveen Prem [1], **Kyle Nash** [1]

[1] University of Alberta

Details

Introduction: Prior research has found that individual differences in neural markers of trait anxiety predict reduced behavioral dishonesty in the Broken Promise paradigm, i.e., a modified trust game with an antecedent promise stage. However, research has not examined if state anxiety, i.e., acute feelings of anxiousness, is also associated with reduced dishonesty. Here, we examined the interaction between trait and manipulated state anxiety on dishonesty along with neural markers of negative arousal, i.e., P300 amplitudes to aversive stimuli, which is sensitive to state anxiety.

Methods: Adult participants (n = 112) first completed a trait anxiety measure, the Behavioral Inhibition System (BIS) scale, before being randomly assigned to control condition or an established state anxiety manipulation that involves reading and summarizing extremely difficult academic text on topic structural

equation modeling that has been found to successfully induce feelings of anxious uncertainty in academic context in the current sample of undergraduate students. They then completed an oddball passive listening task while wearing an EEG headset to measure P300 amplitudes as state anxiety reactions. They finally participated in 9 trials of the broken promise game wherein the participants declared in the antecedent promise stage as a trustee that they would never, sometimes, mostly, or always return half of the money units to the investor. Their subsequent action of keeping or breaking that promise in all trials were computed into a dishonesty score.

Results: Using multiple regression and moderation analyses, we first find that 1) all participants (in both conditions) high in trait anxiety are reliably low in dishonesty, which is consistent with prior studies; 2) however, people low in trait anxiety are higher in dishonesty, but experiencing state anxiety significantly reduces their dishonesty, revealing conditional effect of trait anxiety moderating the association between state anxiety and dishonesty ($\beta = -0.18$, $p < 0.01$); and 3) this effect is further moderated by P300 amplitudes, meaning low trait anxiety individuals in state anxiety manipulation condition with high P300 amplitudes (evidence of increased state anxiety) show the highest decrease in dishonesty as combined conditional effect ($\beta = -0.30$, $p < 0.01$).

Conclusion: These results provide converging neural and psychological evidence that supports state anxiety's association with dishonesty such that low trait anxious humans, who generally do not avoid deception tend to avoid dishonest acts in a state of negative affect or feelings of anxiousness . Given this, we speculate that dishonesty can be muted by transient self-sanctioning emotions.

P1-D-24 - Valence Bias Arises from Both Positive and Negative Responses to Ambiguous Stimuli

Maital Neta [1], **Jordan Pierce** [1], **Nicholas Harp** [2], **James Gross** [3]

[1] University of Nebraska-Lincoln, [2] Yale University, [3] Stanford University

Details

Individuals vary in their tendency to interpret emotionally ambiguous stimuli as relatively more positive or negative, a trait-like difference known as "valence bias." This valence bias has been shown to be stable within individuals over the period of 1-2 years and generalizes to several different categories of emotional ambiguity (faces, scenes, and words). However, this prior research on valence bias has relied upon two-alternative forced-choice (2AFC) valence decisions that conflate the presence of positivity with the absence of negativity, and vice versa. In two preregistered studies (N = 122 and 402, respectively), participants completed a standard 2AFC valence bias task using positive, negative, and ambiguous faces, scenes, and words, and then provided positive and negative unipolar ratings of the same stimuli. Results revealed that the 2AFC valence bias task captured independent effects of both positivity and negativity across all stimulus classes (Study 1: $F(2,72) = 31.86$, $p < .001$, $R^2 = .47$; Study 2: $F(2, 399) = 77.72$, $p < .001$, $R^2 = .28$) with both unipolar positive (Study 1: $b = -15.34$ (95% CI [-20.24, -10.45]), $t = -6.25$, $p < .001$, $r_{a(b.c)} = -.576$; Study 2: $b = -15.72$ (95% CI [-18.66, -12.78]), $t = -10.51$, $p < .001$, $r_{a(b.c)} = -.460$) and unipolar negative (Study 1: $b = 9.14$ (95% CI [3.69, 14.58]), $t = 3.35$, $p = .001$, $r_{a(b.c)} = .287$; Study 2: $b = 6.71$ (95% CI [3.99, 9.41]), $t = 4.87$, $p = .001$, $r_{a(b.c)} = .207$) ratings explaining significant variance in 2AFC valence bias. Individuals with a more negative valence bias showed low unipolar ratings

of positivity and high unipolar ratings of negativity, while those with a more positive 2AFC valence bias showed the opposite pattern. Study 2 also extended these findings to a more racially/ethnically diverse sample with more diverse stimuli. Importantly, emotionally ambiguous stimuli were endorsed as both positive and negative on the unipolar scales, with average ratings falling in the middle of both scales, evidencing their affective value and refuting claims that ambiguity might be "valence-free." The findings also contextualize previously reported relationships between valence bias and psychological well-being, suggesting that the association between more negative valence bias and emotion dysregulation likely reflects greater negativity *and* lesser positivity.

P1-D-25, PS.01.02 - Probability distortions reflect boundary effects under cognitive noise

Saurabh Bedi [1], **Gilles De Hollander** [1], **Christian Ruff** [1]

[1] University of Zürich

Details

Experiments consistently reveal deviations from classical economic theory. One notable example is probability distortions in experiments involving risk. Intriguingly, apparently similar distortions can also be observed in simple perceptual judgments of relative frequency, and more recently in experiments involving "complexity" but not risk. We propose a comprehensive mechanistic explanation for this phenomenon based on a domain-general model of resource-rational cognitive representation of bounded variables. Our model characterizes probability distortions observed in perception, risk, and complexity experiments as central tendency effects within the naturally bounded range of probabilities (0-1), which emerge mechanistically from cognitive noise. Unlike existing Bayesian and efficient coding models, our approach doesn't assume specific prior distributions. We show that this parsimonious model captures both the distortion and variance in responses seen across experiments, and we directly test the causal validity of our model in an experiment on probability distortions in risky choice.

In our experiment, participants (N=63) saw lotteries with varied probabilities but fixed payoffs and reported corresponding certainty equivalents. We used a within-subject 3x2 factorial design with three probability ranges (naturally bounded from 0-1, and artificially bounded from 0-0.5 and 0.5-1) and two cognitive noise levels (complexity of presented probability information).

Our results align with all our model-predicted preregistered hypotheses. Specifically, we observed overestimation of small probabilities (natural boundary at 0) and underestimation of large probabilities (natural boundary at 1) that amplified with increased cognitive noise ($p<2e-16$). Additionally, we found comparable over/underestimation above/below the artificially induced boundary at 0.5 in the high cognitive noise condition ($p<2e-16$) and an interaction effect of the artificial boundary at 0.5 and cognitive noise ($p<2e-16$). Finally, we observed decreased variance in valuation close to the artificial boundary at 0.5 for high cognitive noise ($p<0.001$). This confirms our central model assumptions: The characteristic inverse S-shape pattern of probability distortions causally emerges from boundary effects and cognitive noise.

Our theory predicts distortions and variance in both economic and perceptual experiments involving probabilities. However, we show with model simulations that it is applicable also to other puzzling

choice phenomena, such as stake size effects and overall value effects on variability/accuracy. These model predictions can be tested with causal manipulations in future experiments.

P1-E-26 - Heterogeneity in cognitive capacity and impulsivity predicts the subjective cost of self-control

Nancy (Jiyan) Mao [1], **Kleio Jiang** [2], **Sophia Vranos** [2], **Candace Raio** [1]

[1] New York University, [2] NYU Grossman School of Medicine

Details

BACKGROUND: A growing body of work has demonstrated that the deployment of cognitive control is perceived as subjectively effortful to humans. We recently extended this work to the domain of self-control by measuring the monetary cost choosers were willing to pay in order to use precommitment strategies to avoid tempting rewards that may lead to self-control failures. This work suggests that the extent to which individuals value precommitment can be taken as a proxy for how costly they find exercising self-control to be. Here, we extended this work to examine how heterogeneity in cognitive capacity and impulsivity—two factors widely known to influence goal-directed behavior—shape the perceived cost of self-control.

METHODS: 90 healthy dieters completed a decision-making study in which they first completed a cognitive battery of tasks that included a working-memory measure (Automated Operating Span, or OSPAN) and an intertemporal choice task in which choosers made binary choices between smaller, sooner and larger, later monetary rewards. Participants returned a week later and completed our self-control decision task, during which they viewed images of low, medium and highly-tempting snack foods and reported how much they would be willing to pay to avoid each these food rewards across different amounts of time. OSPAN scores and proportion of immediate (vs. delayed) rewards chosen were used to index working-memory capacity and impulsivity, while average WTP at each level of temptation served as a metric of self-control costs.

RESULTS: Linear mixed-effects models revealed that greater working-memory capacity predicted higher WTP to avoid self-control across each temptation level, suggesting that higher working-memory may lead to a better capacity to prospectively estimate how costly self-control will be once temptations are encountered. In contrast, higher impulsivity predicted lower WTP to avoid temptation, suggesting that self-control costs are estimated to be lower in more impulsive individuals.

CONCLUSION: Our findings provide novel insight into individual variability underlying self-control costs by demonstrating that working-memory and impulsivity have distinct and inverse effects on the

prospective estimation of these costs. This finding suggests that cognitive capacity and impulsivity may be predictive of the extent to which individuals perceive precommitment to be a valuable strategy and point to constructs often related to goal-directed control as impacting how subjective self-control costs are constructed. Future work may seek to determine the extent to which working memory and impulsivity not only predict the estimation of self-control costs but the subsequent success of such strategies.

P1-E-27 - Impatience and present bias do not determine weight loss in obesity: Evidence from lab-in-field and nationally representative data

Agnieszka Tymula [1], **Stefanie Schurer** [1], **Xueting Wang** [2]

[1] University of Sydney, [2] Univeristy of Sydney

Details

The World Health Organization estimates that over 890 million adults and 160 million children and adolescents are currently classified as having obesity. People with obesity face several long-term health risks. Global costs of overweight and obesity are predicted to reach US$ 4 trillion per year by 2035, a cost equivalent to that of the Covid-19 pandemic (World Obesity Federation, 2023). Despite the widely held belief that people should be able to lose weight if they control their behavior, eat less, and exercise more, we lack solid evidence on the association between self-control or patience and success in losing weight which undermines efforts to design successful treatment strategies for obesity.

In this project, we investigated whether people with higher self-control, patience, and risk tolerance are more likely to adhere to a health intervention and are more successful in losing body fat. To answer this question, we use health and economic preferences data of 293 participants who completed a 12-month health intervention. All participants regularly met with a dietician who provided individualized diet and exercise advice. Importantly for our study, during three of these visits – at the start, at six months, and at 12 months, participants' body fat percentage was measured by trained medical staff using the Dual-Energy X-ray Absorptiometry (DXA) scan. Body fat percentage is more useful than Body Mass Index (BMI) in capturing a person's health state because it distinguishes between fat and muscle tissue. Importantly, at the same three appointments, we measured participants' self-control, patience, and risk tolerance using standard incentivized choice tasks. We supplement our analysis with a nationally representative household-based longitudinal survey with over 6,000 adults with similar BMI, education, and gender composition.

Among participants who completed the trial, we found no evidence that self-control, patience, or risk tolerance measured at the start of the health intervention trial predict success in improving body fat composition. This result replicates in the much larger, nationally representative sample. We did however find that participants who are more patient are more likely to complete the trial. Our findings lead us to conclude that increased patience may be a useful trait for weight loss but only through treatment adherence when effective treatments are available.

P1-E-28 - Heterogeneity in state anxiety and depression differentially affects self-control cost estimates

Sophia Vranos [1], **Kleio Jiang** [1], **Nancy Mao** [1], **Candace Raio** [1]

[1] New York University

Details

Background: Self-control failures are common during negative affective states. Existing work has shown that it is cognitively costly to exert self-control. We previously showed that the cognitive cost of self-control can be measured behaviorally and that it increases under stress. However, this work focused on control costs when confronted with temptation. It is unclear how states of anxiety and depression–which are linked to lapses in self-control–may affect the prospective estimation of control costs. Here, we sought to examine how prospectively estimated self-control costs relate to self-reported anxiety and depression.

Methods: Healthy dieters (n=61, M_{age}=30.7) completed a two-session behavioral study. During session 1, participants provided self-reported foods ratings and completed a state (STAI-S) and trait anxiety inventory (STAI-T) and the Beck Depression Inventory (BDI-II). A week later, participants repeated the STAI-S and completed the self-control decision task. On each trial, participants viewed a food image that varied on temptation, quantity and exposure time with the food and reported trial-by-trial willingness-to-pay (WTP; from a study endowment) to avoid the food depicted on each trial. A realization phase followed, where one trial was randomly selected and played out in a standard economic auction procedure. We examined how choice behavior (bids) related to self-reported measures of anxiety and depression using linear mixed-effect models in R.

Results: Participants with higher state anxiety bid more to avoid using self-control (p=.023), with higher levels of temptation revealing stronger associations between anxiety and bids (anxiety X temptation interaction; p<.001), while those with higher depression bid less to avoid using control.
We also observed an interaction between anxiety and depression (p=.022), such that the association between state anxiety and bids was positive for participants with lower depression scores
but changed to negative for those with higher depression scores. Thus, those with higher anxiety were likely to bid more, unless they also had higher depression scores.

Conclusion: Our study reveals a novel relation between anxiety and WTP to avoid using self-control, with this relationship becoming stronger as temptation increases. It also reveals how higher depression scores reverse this relationship, leading to lower WTP to avoid control. We propose that in the same way anxiety allows one to anticipate threats, it may also allow one to anticipate self-control failures, leading to higher WTP to avoid control. However, a more motivation-based pathology like depression may not be positioned to exert the same effect.

P1-E-29, PS.01.03 - Serotonin increases willingness to wait for delayed rewards in humans

Karolina Lempert [1], **Laura Zaneski** [2], **Arjun Ramakrishnan** [3], **Daniel Wolf** [2], **Joe Kable** [2]

[1] Adelphi University, [2] University of Pennsylvania, [3] Indian Institute of Technology Kanpur

Details

Objective: People often have to decide how long to wait for rewards that will arrive at an uncertain time in the future. Previous work has suggested that people approach these willingness-to-wait decisions by trading off the benefits of continued waiting against opportunity costs of waiting, the balance of which can change as they wait. Thus, willingness-to-wait decisions might share neural mechanisms with foraging decisions, which also involve an ongoing comparison between a current resource and possible alternatives. Dopamine and serotonin have been proposed to play opposing roles in foraging, with dopamine promoting exploration and serotonin promoting exploitation. Here we investigated their roles in willingness to wait, in a pre-registered, crossover, double-blind, placebo-controlled study in young adults. We examined how increasing dopamine (with 200 mg levodopa) and increasing serotonin (with 20 mg escitalopram) affected willingness to wait. Based on their proposed roles in foraging, we expected that levodopa would decrease willingness to wait and escitalopram would increase it. Methods: Participants (n = 42) completed three sessions, each one week apart. At each session, they received either placebo, escitalopram, or levodopa (counterbalanced order), and then performed decision-making tasks, including a willingness-to-wait task. On each trial of this 10-minute task, they waited through a random delay (0.2-40 s) for a 10¢ reward. They did not know how long the 10¢ would take to arrive, and they could quit waiting at any time and go to a new trial. Reward delays were distributed such that there was no optimal time at which to give up waiting. We plotted the probability that participants were still waiting for the reward at any given time point within the 40 s delay period; the area under this curve (AUC) was our measure of willingness to wait, with higher values indicating longer wait times before quitting. Results: We first decided on the inclusion of covariates according to our pre-registered plan, before running two mixed-effects regressions (one for levodopa, one for escitalopram) to test for the effects of drug on AUC. After controlling for anxiety levels and condition order, there was a significant effect of taking escitalopram on AUC, such that people were more willing to wait for delayed rewards after taking the serotonergic drug (b = 3.93; p = 0.012). There was no effect of levodopa on willingness to wait (b = -0.64; p = 0.398), however. Conclusions: This result nicely complements the rodent literature, which has shown a critical role for serotonin in waiting for rewards that are delayed by seconds. Possible mechanisms will be discussed.

P1-E-30 - Computing Confirmation Bias in Impression Formation: A Comparison between Social and Non-Social Contexts

Yingjie Li [1], **Jean Daunizeau** [2], **Fred Hasselman** [1], **Johan Karremans** [1], **Sabine Geurts** [1]

[1] Radboud University, [2] Paris Brain Institute

Details

Study Objectives:

Impression formation is a dynamic process involving the integration of information over time, yet the precise mechanism remains unclear. This study aimed to explore how individuals form impressions of

both social targets (employees) and non-social targets (vacation houses), focusing specifically on confirmation bias. Computational models were developed and tested to explain this process.

Methods:

The study employed a 2 (first impression: positive vs. negative) x 2 (target type: social vs. non-social) mixed design. Over six days, 86 participants received daily information about a virtual employee and a vacation house. They were divided into two groups, receiving either increasingly positive employee behaviors and increasingly negative vacation house information, or vice versa. Participants were requested daily to make hypothetical decisions regarding employee contracts and vacation house rentals.

5 Bayesian computational models, incorporating implicit confirmation bias strategies such as over-weighing confirmatory information and under-weighing contradictory information, were developed and compared against the data. The model parameters were configured so that 0 indicated no bias, positive numbers indicated confirmation bias and negative numbers implied anti-confirmation bias. Model selection was based on the best fit concerning both the initial impression and the impression target. Furthermore, we compared the strength of the confirmation bias (after weighted averaging across all models) under different conditions (i.e. first impression and the impression target).

Results:

The model selection showed no difference in strategies based on the type of target. This is because the same computational model was consistently chosen for both social and non-social targets. Nonetheless, distinct strategies did emerge depending on initial impressions. Participants demonstrated a tendency to distort information opposing their initial impression, either by overweighing or underweighing it.

Parameters averaged across the models confirmed the presence of confirmation bias ($F(4,168)=48.87$, $p<.001$, $R^2=0.538$). Specifically, confirmation bias is stronger ($F(1,168)=35.10$, $p<.001$, $R^2=0.173$) when the initial impression is positive($M=0.54$, $SD=0.52$) compared to negative ($M=0.19$, $SD=0.28$). Participants exhibited a (slightly) stronger confirmation bias($F(1,168)=4.01$,$p=.047$, $R^2=0.023$) toward social targets ($M=0.47$, $SD=0.51$) compared to non-social ones($M=0.26$, $SD=0.36$).

Conclusions:

This study demonstrates the existence and potential computational mechanisms of confirmation bias in a real-world context, shedding light on the dynamics of impression formation.

P1-F-32 - Mechanisms of Learning in social networks: one-shot averaging or error-driven gradual updating

Julien Benistant [1, 2], **Alireza Soltani** [3], **Fréderic Moisan** [4], **Jean-Claude Dreher** [1]

[1] National Center for Scientific Research, [2] Univ. Lille, CNRS,, [3] Dartmouth College, [4] Université Claude Bernard Lyon 1

Details

We acquire a substantial amount of our knowledge through observing friends, relatives, and colleagues within our social network, a phenomenon referred to as social learning. One of the most successful models of social learning, the DeGroot model, posits that we adjust our beliefs based on the average opinions of our social connections. In contrast, reinforcement learning (RL), assumes learning is sequential and error-driven, based on the difference between expectations and outcomes. Although both models are implicated in spreading information in social networks, it is unclear which model the brain's computations conform to. To answer this question, we used these two models and their variants to fit data from two recent studies on information spread in social networks. In both studies, participants performed an observation-learning task where they were assigned to different nodes in a network and had to determine the state of the world underlying their observations, either through private signals or by observing others' guesses. We found that the DeGroot model better captured learning behavior, despite variations in learning strategies across tasks and networks. Moreover, our theoretical account and simulations of behavior in the task demonstrated that inconsistencies between the DeGroot and RL models lead to significantly different outcomes in terms of global consensus and accuracy, depending on the network structure. Overall, our results provide new insights into mechanisms of information propagation in social networks and how underlying computations can be detected in future experiments.

P1-F-33 - Neural responses to social rejection reflect learning about relational value and reward

Begum Babur [1], **Chelsey Pan** [1], **Leor Hackel** [1], **Yuan Chang Leong** [2]

[1] University of Southern California, [2] University of Chicago

Details

Social rejection hurts, but it can also be informative: through experiences of rejection and acceptance, people identify partners interested in connecting with them and choose which ties to cement or to sever. Past work has examined how the brain responds to the hurt of rejection, identifying a social rejection network that overlaps with responses to pain. Yet, these regions also respond when people update their beliefs, suggesting these regions may underpin learning from social experience. We used computational neuroimaging to ask whether neural responses reflect two potential kinds of social learning. First, people generally learn from rewarding outcomes, which may include concrete opportunities for interaction. Second, people track the "relational value" others ascribe to them – an internal model of how much they think others value them. To dissociate these forms of learning, participants in a neuroimaging study (N=40) repeatedly tried to match with others in a social game. Feedback revealed whether they successfully matched (a rewarding outcome) and how much the other person wanted to play with them indicated by ranks (relational value). A Bayesian cognitive model including both relational value and reward value fit the data better than models including one strategy or the other (protected exceedance probability=1). Using fMRI we dissociated the two different learning mechanisms: receiving a better outcome than expected (positive outcome prediction errors) was linked to brain regions involved in reward-based reinforcement (ventral striatum), whereas receiving a

different rank than expected (unsigned rank prediction errors) was linked to regions in the social rejection network (e.g.,dorsal and ventral anterior cingulate cortex, anterior insula). This suggests that when participants received a different rank than expected, they updated their beliefs about what they think their interaction partner is feeling towards them. If these regions encode participants' internal model of relational value, then voxel patterns in these regions should be similar when participants view players who valued them to a similar degree. We used participants' post-task ratings of how much they think each player liked them as their subjective perceptions of relational value. Using representational similarity analysis (RSA) we found that players who elicited more similar voxel patterns were rated more similarly (M=.13, SD=.264, t(39)=3.08, p=.004). Here, we identify precise computations underlying brain responses to rejection and support a neurocomputational model of social affiliation in which people build an internal model of relational value and learn from rewarding outcomes.

P1-F-34 - Experience and advice consequences shape information sharing strategies

Hernan Anllo [1], **Uri Hertz** [2], **Stefano Palminteri** [3], **Nichola Raihani** [4], **Gil Salamander** [2]

[1] École Normale Supérieure-PSL University, [2] University of Haifa, [3] École Normale Supérieure, [4] University of Auckland

Details

Introduction: Individuals often rely on external advice to reduce uncertainty and maximize success during learning. However, advisers are often themselves individuals who continuously learn. Here, we seek to uncover how advising behaviors transform throughout this learning process.

Methods**:** We designed a 2-armed bandit task embedded in a classic reinforcement learning framework. After each decision, but before decision outcome, participants had the option of broadcasting their choice as advice to future players. In our first study, participants could share advice at will. In our second study we analyzed the impact of external consequences on the decision of advising: we conducted four experiments introducing respectively (1) a monetary fee after giving advice, (2) advise being submitted to a public ranked scoreboard, (3) a warning to the adviser that future participants would forcibly rely on their advice, (4) a replication of the free advice condition.

Results**:** Our first study (N=373) indicates that participants exhibited a general bias towards sharing advice early on. However, as advisers accrued knowledge on the value of options, willingness to share advice and advice quality both increased (P(advice) ~ Trial number: Trial number main effect: $\chi2(1) = 109$, $p < 0.001$). Model estimations of choice probability confirmed that the tendency to share advice increased as advisers gradually moved from exploratory behavior, characterized by trial-and-error decision-making, towards more exploitative behaviors (choice probability estimation: $p(Choicet) = e\beta \cdot Q(\text{Chosen})t / (e\beta \cdot Q(\text{Chosen})t + e\beta \cdot Q(\text{Unchosen})t)$; P(advice) ~ P(choice): coefficient = 2.99(2.82, 3.16), $\chi2(1) = 1206$, $p < 0.001$). In the second study (N=210) monetary losses led to a decrease in participants' advising tendencies but not advice quality, and the promise of a public scoreboard incremented overall sharing behavior (mean advice(SEM): No cost [replication] = 0.60(0.01), Money = 0.27(0.01), Scoreboard = 0.76(0.01), Warning = 0.64 (0.01); Advice consequence main effect: $\chi2(3) = 56$, $p < 0.001$).

Conclusions**:** The results highlight that advisers tend to share informative choices, rooted in their own experience. Yet, while this informative core remains integral, for the most part the exact quality and frequency of advice is permeable to the external consequences associated with advice-giving. Our findings emphasize the interrelation between individual learning, external incentives, and psychological predispositions. We posit that these insights could be pivotal for developing models that predict and explain human behavior in peer-to-peer learning platforms.

P1-F-35 - Mapping the cognitive and neural processes of inferring others' preferences from observing how they decide

Sophie Bavard [1], **Sebastian Gluth** [1]

[1] University of Hamburg

Details

Objective
While substantial research has been conducted on how humans infer hidden preferences of others from observing their choices, the effect of their response times (RT) on such inferences has received little attention. We recently investigated this question and showed that humans can infer other people's social preference, even when observing their RT alone. In parallel, a recent EEG study identified a brain signal which encodes the mismatch between the expected and actual length of elapsed time. Here, we investigate whether this neural marker of time estimation is also relevant when judging the speed of a decision and thus involved in inferring other people's preferences from RT.
Methods
We designed a novel experimental setup featuring a 2x2 factorial design which independently varied perceived time window and decision difficulty. N = 50 participants observed someone else's decisions in a Dictator Game while being recorded with EEG. Trials were divided into two blocks. First, participants observed decisions categorized strictly as easy/fast or hard/slow in the learning phase, to establish baseline inference of preferences. Second, in the ensuing post-learning phase, we introduced unexpected combinations of speed and difficulty (i.e., easy/slow and hard/fast) by fitting and simulating a drift-diffusion model to the Dictator Game data. Finally, we implemented a replication of the time estimation task.
Results
Our findings successfully replicate (1) our previous work showing that observers can learn other people's social preferences when observing their RT alone, and (2) the established EEG results in the time estimation task. During the learning phase, we find the expected trial type effect, with a significant difference in the amplitude of the fronto-central EEG signal between easy/fast and hard/slow observed trials. Critically, in the post-learning phase, we find that the targeted EEG marker indeed dissociates between both decision speed and difficulty, suggesting it to contribute to people's inferences about others' social preferences.
Conclusions
Our research marks the first use of EEG to explore how decision RT alone influences the perception of others' preferences. We identify a brain signal whose amplitude depends on changes in time perception and decision difficulty. These insights enhance our understanding of the neural mechanisms

underpinning social perception and offer promising directions for future research into how humans interpret and predict the actions of others based on minimal cues like decision speed.

P1-F-36 - Egocentric anchoring-and-adjustment of social knowledge in the hippocampus

Marta Rodriguez Aramendía [1], **Mariachiara Esposito** [1], **Raphael Kaplan** [1]

[1] Universitat Jaume I

Details

Growing evidence suggests the hippocampus represents abstract relational knowledge, including social information, similar to how physical locations are represented on a map. Yet how map-like representations of abstract knowledge are influenced by personal biases is unclear. We test whether a prominent egocentric bias involving an implicit reliance on self-knowledge when rating others, anchoring-and-adjustment, affects how the relative attributes of different social entities are learned. Participants (N=20) provided likelihood ratings of partaking in everyday activities for themselves, fictitious individuals, and familiar social groups. Subsequently during functional neuroimaging, participants learned a stranger's preference for an activity relative to one of the fictitious individuals and decided in a two-alternative forced choice(2AFC) how the stranger's preference related to the groups' preferences. Egocentric anchoring-and-adjustment was present when participants rated the other entities (b=0.21, t(19)=2.18, p=0.030), where anchoring interfered with performance when compared groups weren't too different or similar to self ratings. Hippocampal signals similarly related to group-self rating discrepancy, suggesting that hippocampal sensitivity to egocentric anchoring influences dynamic world-centered inferences. Linking the hippocampus to individual subjective biases, increased similarity in hippocampal signal patterns over trials reflected participants' propensity towards egocentrism. These findings imply that personal preferences help shape the hippocampus' mnemonic representation of other people's preferences.

P1-F-37 - The blessing and curse of value-shaping imitation

Isabelle Hoxha [1], **Stefano_ Palminteri_** [1]

[1] École Normale Supérieure

Details

Recent empirical studies suggest that in the context of social reinforcement learning, imitation is instantiated as a form of value shaping process (VS), where the actions of another player (the demonstrator) affect the value function of the learner[1]. This finding is in contrast with a previous hypothesis, Decision Biasing (DB), whereby the demonstrator's action is taken into account in the decision policy of the learner, instead of their value function[1,2]. Recent studies also show that, consistently with evolutionary theory, imitation is controlled through a meta-learning process, where inaccurate demonstrators are progressively less imitated[1,3].

In the present study we aimed at grounding value shaping imitation on a normative status.

To do so, we simulated agent's behaviours using different imitation strategies in increasingly complex scenarios and compared the obtained learning curves and accuracy to classical reinforcement learning outcomes (i.e. in the absence of imitation) and vicarious learning, i.e., a social learning situation where the player learns not only from the demonstrator's choices but also from their outcomes.

We showed that, while in simple settings VS imitation generally outperforms other forms of imitations and no imitation. Notably, VS leads to better performances in difficult environments (i.e. when the difference between the options' values is small) and is more resilient to different phenotypes of the demonstrator (i.e. across their skill levels).

However, VS is maladaptive in more complex settings. More specifically, when imitation is bidirectional and the environment unstable, VS leads to the worst possible performance. This catastrophic performance is enhanced by a meta-learning process, in particular when meta-learning is implemented on the basis of the similarity between the two agents, and we speculate that it may be contributing to echo chambers and opinion polarization in real life.

Finally, we introduce two modifications of the VS model to overcome these difficulties: Conditional Value Shaping, and meta-learning on the basis of the validity of the demonstrator.

We show that both models are more resilient to volatile environments and lead to more robust performance across task difficulties and demonstrator's skill levels.

1. Najar et al., PLoS Biology (2020)
2. Burke et al., PNAS (2010)
3. Selbing et al., Sci. Rep (2017)

P1-F-39 - Human cognitive biases extends to interaction with Artificial Intelligence systems

Moshe Glickman [1], **Tali Sharot** [1]

[1] University College London

Details

Objective: In various domains, humans are increasingly interacting with artificial intelligence (AI) algorithms. From Large Language Models such as ChatPGT to AI applications used to interpret medical imaging. AI systems have become an inherent part of our daily lives. A fundamental question that arises is whether human-AI interactions mirror those with other humans or possess unique characteristics? Here, we investigated whether cognitive biases observed in human-to-human interactions are similarly manifested in human-AI interactions. In particular, we explored whether the "better than average" effect, where individuals tend to perceive their performance as superior to others, generalizes to interactions with AI systems.

Methods: Participants (*N*=379) performed a face averaging task, classifying an array of emotional faces as 'more happy' or 'more sad'. They interacted with an AI system (Convolutional Neural Network) or with human partners. The input and label of the partner was manipulated, creating four distinct scenarios: AI presented as AI, AI presented as human, human presented as human and human presented as AI. In each trial, participants classified the arrays and reported their confidence. Then, they were presented with the response and confidence of their partners, and were asked whether they would like to change their initial decision or not. Finally, participants assessed their own accuracy and that of the partners they interacted with.

Results: Although participants were less accurate than the AI partner ($t(378)=25.2$, $p<0.001$, $d=1.78$) and exhibited similar accuracy to their human partners ($t(378)=0.11$, $p=0.90$, $d=0.01$), they perceived themselves as more accurate than their partners across all conditions: AI presented as AI or as human, and human presented as human or as AI (all p's<0.009). Moreover, we examined the impact of the input (AI vs. human) and label (AI vs. human) on the participants' perceptions of their partners' accuracy. The results revealed a significant effect for both factors: When the partner was labeled as an AI, participants perceived the partner's responses as more accurate ($b=0.13$, $p<0.001$). However, when the actual input came from an AI, participants perceived their partner's responses as less accurate ($b=-0.39$, $p<0.001$).

Conclusions: The findings demonstrate that human cognitive biases, such as the "better than average" effect, extends to interactions with AI systems, potentially resulting in suboptimal decisions. Furthermore, the findings suggest that to achieve better performance, the design of AI systems should take into account humans' cognitive biases that may be expressed in interaction with AI systems.

P1-F-40 - Steeper social discounting after human basolateral amygdala damage

Tobias Kalenscher [1], **Luca Lüpken** [1], **David Terburg** [2], **Jack Van Honk** [3]

[1] Heinrich Heine University Düsseldorf, [2] Utrecht University, [3] University of Cape Town

Details

The amygdala, traditionally associated with emotions, has recently been linked to social decision-making. Evidence in the animal and human literature suggests that amygdala damage impairs prosocial choice, and that amygdala computes some of the core processes underlying prosociality. However, human patients with focal lesions in the basolateral amygdala (BLA) make moral and economic decisions indicative of hyper- rather than hypoaltruism, so the role of amygdala in prosocial decision-making remains unclear. Here, we hypothesized that one reason for the discrepancy in results could be that the social distance between actor and recipient of a prosocial deed is typically not controlled for. People accept costs to help socially close others, but their generosity declines with social distance, a phenomenon termed social discounting. We predicted that the relation between social distance and generosity depends on BLA. We tested five participants with the rare Urbach-Wiethe disease who had selective, bilateral BLA damage and compared their social discounting behaviour to that of 16 healthy control participants matched for (neuro-)psychological traits and cultural and socioeconomic background. Participants assigned people from their social environment to 8 social distance levels ranging from 1 (the emotionally closest person) to 100 (random stranger on the street). Then, for each person on each social distance, they played a dictator game in which they could share 200 ZAR (approx.

US$ 10) with the respective other person. We computed model-free and model-based markers of social discounting (model-free: area-under-the-curve, AUC; model-based: estimation of hyperbolic social discount parameters). Patients showed markedly steeper social discounting than controls (AUC: two-sided t-test t_{19}=2.891, p<.01, Cohen's d=.875; discount parameter log(k): four out of five patients exceeded the upper limit of the 95% confidence interval; the fifth patient had overall lower generosity levels than controls), indicating a sharper decline in generosity across social distance. This difference in social discounting between patients and controls could not be explained by empathy or personality traits. The observation of steeper discounting in BLA-lesioned patients likely reflects their tendency to resolve the social-distance-dependent trade-off between self- and other-regarding motives in favour of the former; consistent with recent evidence that the human BLA supports model-based decision-making. Our findings expand our understanding of the computational mechanisms in the amygdala that regulate prosocial choice.

P1-F-41 - Value-based preferences predict views on redistribution

Elijah Galvan [1], **Alan Sanfey** [1]

[1] Donders Institute

Details

Economic inequality is one of the most pressing threats to political stability in the twenty-first century, yet efforts to reduce the widening gap between the rich and the poor often receive little public support – even amongst those who would financially benefit. Although much research has sought to probe the motives responsible for the proliferation of these behaviors, there are conflicting accounts of the psychological processes responsible for this paradox of redistribution. It is important, therefore, to understand what leads to the willingness, or not, to reallocate resources, and what contextual factors influence these preferences. To address this, we developed a novel experimental task to explore redistributive decision-making. In a large sample of participants (N = 231), we found several distinct strategies that characterized redistributive choices – namely equity, equality, and payout-maximization. Specifically, we find evidence of computationally tractable redistributive strategies, and suggest that the adoption of such strategies is sensitive to both individual factors and the context of the inequality itself – that is, how it has originally occurred. Further, these strategies may guide the formation of political opinions: strategy selection in our experimental task predicted specific relationships between participants' policy views on redistribution and how they would benefit from it, independent of perceptions about the cause of inequality. Individual differences in strategy preferences determine redistributive preferences, rather than differing perceptions about what causes inequality.

P1-F-42 - The Impact of Self-Esteem and Status Disclosure on Conspicuous Consumption: An fMRI Study

Saim Kim [1], **Jinhee Kim** [1], **Gahyun Lim** [1], **Hackjin Kim** [1]

[1] Korea University

Details

Objective: This study investigated the impact of self-esteem on conspicuous consumption in varying status disclosure contexts and identified associated neural mechanisms.

Methods: Male participants (n=47, mean age =23.83) were recruited and led to believe their virtual characters would be evaluated as potential team members by females. Participants completed a cognitive task with two others, and were ranked based on their performances, unknowingly being ranked 2nd and earning equal amounts. Using these incentives, participants purchased outfits for their virtual characters during a 2 x 2 (competitor status: higher, lower; disclosure: shown, hidden) decision task inside an fMRI scanner. Twenty brand options, ranging from mass-market to luxury brands, were selected and priced at 100-500 Korean Won based on an independent survey.

Results: A repeated-measures ANCOVA with self-esteem as a covariate revealed a significant 3-way interaction of disclosure, competitor status, and self-esteem ($F(1,45) = 7.43$, $p < 0.01$). Lower self-esteem participants showed a significant main effect of disclosure (hidden > shown, $F(1,22) = 8.17$, $p < 0.01$) and increased in mean purchase prices when the competitor's status was higher and hidden (higher hidden > higher shown, $t(22) = 2.98$, $p < 0.01$). Higher self-esteem participants showed no significant changes across conditions.

fMRI analysis (n=38, excluding 9 for excessive head motion or repetitive choices) using multiple regression analysis with self-esteem on the decision interaction contrast (higher [shown – hidden] – lower [shown – hidden]) revealed significant clusters in the pgACC/rmPFC (MNI [10, 42, -6], FWE cluster-level corrected). Post hoc median split analysis of peak voxel beta confirmed behavioral patterns. That is, lower self-esteem participants showed a significant interaction of disclosure and status ($F(1,15) = 12.845$, $p < 0.01$) and significant increases in purchase price in the higher hidden condition (higher hidden > higher shown, $t(15) = 2.36$, $p = 0.032$), whereas higher self-esteem participants showed no differences in pgACC/rmPFC activity across any pairs of conditions.

Conclusions: Individuals with lower self-esteem show increased conspicuous consumption, especially when competing with higher-status individuals whose status is hidden. The pgACC/rmPFC, associated with self-enhancement under social observation (Izuma et al., 2010; Yoon et al., 2021), encodes individual differences in the subjective need for self-enhancement through conspicuous consumption.

P1-F-43 - Neuroaffective mechanisms of interpersonal trust vary across cultures

Huan Wang [1], **Yijie Fang** [2], **Christy Wang** [1], **Brian Knutson** [1], **Jeanne Tsai** [1]

[1] Stanford University, [2] Fudan University

Details

Extending trust to another involves pursuing gains related to cooperation, but also avoiding losses related to betrayal. Pursuing interpersonal gains has been associated with increased activity in mesolimbic regions including the Medial PreFrontal Cortex (MPFC) and Nucleus Accumbens (NAcc),

which have been implicated in benefit-cost analyses and positive arousal, respectively. Avoiding interpersonal losses has been associated with increased Anterior Insula (AIns) activity and linked to negative arousal. Behavioral research suggests that overall, East Asians trust strangers less than European Americans, which might be related to cultural differences in viewing social relationships in ranked hierarchy. How neural mechanisms of trust vary across cultures has received less investigation. To address this, across two studies, we first assessed brain activity using Functional Magnetic Resonance Imaging (FMRI) in 25 healthy European Americans and 27 Chinese adults as they played 72 trials of a single-shot Trust Game. On each trial, subjects saw picture of a partner's face, followed by information about their trustworthiness (i.e., rated by previous players based on partner's reciprocation history), and finally a prompt to invest (out of a $6 endowment), which served as a proxy for trust. We then corroborated the FMRI results using 36 trials of affective ratings on the same partners in samples of 112 European Americans and 112 Chinese. Behaviorally, European Americans invested in strangers more than Chinese ($F=20.10$, $p<.001$). Neurally, pre-choice AIns activity negatively predicted choices to invest for European Americans ($B=-.56$, $p=.007$), whereas pre-choice MPFC ($B=.32$, $p<.001$) and NAcc ($B=.36$, $p=.047$) positively predicted choices to invest for Chinese. Further, European Americans rated untrustworthy partners with higher negative arousal than Chinese ($t=2.10$, $p=.037$). Finally, across both FMRI and behavioral samples, valuing social hierarchy negatively predicted trust across cultures ($r=-.32$, $r=-.18$, $ps<.05$). These findings suggest that cultural values shape the default levels of trust towards strangers, which may then influence the motivations to trust. On the one hand, European Americans with higher trust of strangers were more motivated to avoid betrayal and so distrusted bad partners. On the other hand, Chinese with lower trust of strangers were more motivated by potential cooperative gain, and so trusted good partners. Together, these findings highlight the intersection of cultural expectations and reputation in the choice of whom to trust, which may have implications for improving cross-cultural communication and cooperation.

P1-G-44 - Sensitivity to cognitive rewards changes through development

Tal Nahari* [1], **India Pinhorn** *[1], **Tali Sharot** [1]

[1] University College London

Details

Objective: From play in infancy to solving crossword puzzles at eighty, humans find fascination and pleasure in cognitive activities across the lifespan. We propose that such cognitive activities are inherently pleasurable because they lead to specific mental outcomes, which we call 'cognitive rewards'. Understanding cognitive rewards is crucial, as the link between cognition and pleasure likely plays a significant role in childrens' learning progress. Here, we test the hypothesis that sensitivity to diverse cognitive rewards (agency, novelty, uncertainty reduction, epistemic knowledge) increases across the developmental trajectory.

Methods: Participants (Exp 1 N = 213, age 4-12) played a series of 'games' which tested sensitivity to putative cognitive rewards (agency, novelty, uncertainty reduction, epistemic knowledge) compared to neutral paired alternatives. We collected two canonical behavioral signatures of reward – liking and

wanting. Participants rated how much they liked each game and selected between each putatively rewarding game and its paired alternative. We then examined how sensitivity to different cognitive rewards changes with age.

Results: all of our categories were significantly different when comparing the rewards to alternatives (mean differences ranging between 0.6 to 1.05 on a 7-item scale, all p<.001). We performed a separate dimensionality reduction (PCA) across the liking and wanting responses to cognitive rewards, which allowed us to extract sensitivity scores for each participant on different types of cognitive reward. We saw that the first dimension accounts for 34% of the variance, loading on all of the cognitive rewards, and these were significantly correlated with age (liking: r_{213}=0.32, p<.001; wanting: r_{213}=0.45, p<.001), showing the hypothesized increased sensitivity to cognitive rewards with age. Moreover, we used a linear regression to test whether the increase of reward sensitivity with age is driven by specific categories, and found that those are significantly driven by epistemic knowelede (b=1.13, t = 4.7, p<.001) and novelty (b=1.3, t=4.53, p<.001).

Conclusion: These findings provide new insights into the developmental trajectory of cognitive reward sensitivity, with important implications for the study of human cognition and the improvement of educational systems.

P1-G-45 - Effects of depression and self-esteem on self-related learning and attributions of outcomes

Annalina Mayer [1], **Alexander Schröder** [1], **Nora Czekalla** [1], **David Stolz** [1], **Sören Krach** [1], **Tobias Kube** [2], **Laura Müller-Pinzler** [1]

[1] University of Lübeck, [2] University of Koblenz and Landau

Details

Objective: People vary in how they explain the outcomes of their own behavior. For example, while most healthy people attribute successes to internal causes (e.g., their own abilities) and failures to external causes (e.g., bad luck), depressed individuals show the opposite pattern. Since so-called attributional styles have mostly been treated as interindividual traits in past research, little is known about how causal attributions arise in the moment and influence how people form beliefs about themselves and the world. Recent studies in neuroscience and psychology have started to examine the effects of momentary causal attributions on learning about the external world. The goal of this study was to examine attributions of successes and failures in *self-related* learning and their modulations by depression and self-esteem.

Methods: Fifty-five university students completed a self-related learning task. On a trial-by-trial basis, they were asked about expectations on their performance and received feedback and in four different

task domains. By using controlled feedback sequences, participants were led to believe that they performed well in two domains and poorly in the other two domains. Additionally, they were informed that in two domains, their feedback might sometimes be manipulated by the computer, allowing different causal attributions of the received feedback. Depressive symptoms and self-esteem were assessed using validated questionnaires.

Results: Participants updated their expectations more strongly in the condition in which they believed the feedback was related to their true performance, compared to when they believed the computer could intervene. In the condition in which they received predominantly good feedback, participants on average attributed the feedback more often to their own performance (vs. intervention by the computer) compared to the condition when they performed poorly. Participants with stronger depressive symptoms updated more negatively across all conditions, leading to overall worse performance expectations. Participants with lower self-esteem updated more negatively specifically in the condition where they received predominantly good feedback. Self-esteem, but not depression, also moderated the effects of feedback condition (good vs. bad) on causal attributions of the received feedback.

Conclusion: We provide evidence that self-esteem and depression impact momentary causal attributions of behavioral outcomes and their integration into self-related beliefs. This might have important implications for social decision making.

P1-G-46 - What do your eyes smell? A multimodal study of visual consumer decision making of perfume bottles.

Bernardo Canelas [1], **Catarina Ribeiro** [2], **Maria Vânia Nunes** [3], **Hugo Ferreira** [4]

[1] Instituto Superior de Psicologia Aplicada, [2] UCP - Universidade Católica Portuguesa Institute of Health Sciences, [3] Center for Interdisciplinary Research in Health - CIIS, [4] IBEB - Institute of Biophysics and Biomedical Engineering

Details

Understanding the neurobehavioral and psychological factors that influence consumer decision-making is crucial for gaining insights into human behavior within a broader ecological context. This study aimed to elucidate the interplay between personality, neural activity, and consumer decision-making.

Forty neurotypical participants (23 females, 17 males; mean age = 22 years) were recruited for the study. A multimodal approach was employed, combining eye-tracking, heat map generation, and pupillometry metrics with electroencephalography (EEG) through event-related potential (ERP) analysis. Additionally, individual differences in personality were assessed using the NEO PI-R. This methodology allowed for the investigation of how contextual sensory aspects (e.g., color and shape) and Big 5 personality traits influence preference-based decision-making. Participants were tasked with choosing whether to buy or not buy stimuli consisting of AI-generated images of perfume bottles, designed to represent the market without brand elements but with similar colors and shapes. This aimed to provide

a clearer understanding of decision-making processes by minimizing the influence of previous brand-related emotional experiences.

Preliminary regression analysis revealed that openness to experience, particularly the aesthetics facet, was a significant predictor of the number of decisions to buy, explaining 20.4% of the variance ($F(1, 38) = 9.745$; $p = 0.003$, $R^2 = 0.204$). These decisions were influenced by the shape and color of the perfume bottles presented ($\chi^2(120) = 3034.795$; $p < .001$), with square shapes and transparent flasks being preferred.

Moreover, ERP analysis showed distinct frontal and parietal activity in the P200 and P300 components, with significant differences between participants with low and high openness to experience during positive purchase decision events ($p < 0.05$).

These findings highlight the importance of considering the interplay between individual differences in personality and neural processes in understanding consumer behavior and preference heterogeneity. Specifically, openness to experience and aesthetic sensitivity reflect a functional dissociation in neural processes that increase the number of purchase decisions. By providing insights into these mechanisms, this study contributes to the broader understanding of decision-making processes and their neural underpinnings, with implications for fields such as neuroeconomics and consumer neuroscience, as well as practical applications in neuromarketing and UX/UI design.

P1-H-47 - The importance of timing for learning and motivation

Giorgia Bergmann [1], **Hui-Kuan Chung** [2], **Philippe Tobler** [2]

[1] University of Zurich, [2] University of Zürich

Details

When conditioned stimuli predict the occurrence of an appetitive or aversive outcomes, they acquire the valence of the outcome (Baeyens et al., 1992). However, such forward associations constitute only one half of the associative learning process. Albeit largely overlooked in scientific inquiry, cues associated with the termination of valenced events might be as important as cues that predict their occurrence. Indeed, animal evidence suggests that termination-related cues can acquire valence opposite to the associated outcome itself (timing-dependent valence reversal; Gerber et al., 2019; Yarali et al., 2008). Here, we investigate how backward associated cues acquire value and how they influence motivated behavior in humans.

We use a variant of a Pavlovian-Instrumental Transfer (PIT) task, which consists of three phases. In the Pavlovian phase, subjects learn to associate different cues with the occurrence or termination of rewards or punishments. In the subsequent instrumental phase, subjects perform an action to receive a reward. Finally, in the transfer phase, subjects freely perform the instrumental action while cues are presented one by one. The change in press rate during these presentations relative to baseline reflects the motivational impact of learned cues. In our paradigm, cues are either forward or backward associated with one of three possible outcomes (Reward, Punishment, or Neutral) in the Pavlovian

phase. In addition to button pressing, we use an explicit measure (liking of cues, corresponding to evaluative conditioning).

We present results from a pilot study (N=19) in which we found partial evidence for timing-dependent valence reversal. While a cue predicting reward occurrence increased button pressing in the transfer phase, a cue associated with reward termination decreased it. By contrast, cues associated with punishment occurrence or termination both decreased button pressing. The opposite pattern was evident in the explicit measure. Liking showed timing-dependent valence reversal for punishment but not for reward. Thus, subjects explicitly liked the cue associated with punishment termination more and the cue associated with occurrence of punishment less after learning. Conversely, cues forward or backward associated with reward were both liked more at the end of the task.

Together, we provide preliminary evidence for domain- and measure-specific timing-dependent valence reversal. Our findings corroborate a previous report (Andreatta et al. 2010) of a rift between implicit and explicit measures in backward associations with aversive outcomes. In future research, we will assess the current findings in larger samples.

P1-H-49 - Confirmation Bias during Formation and Revision of Self-Beliefs: The Role of Expectation and Confidence

Alexander Schröder [1], **Nora Czekalla** [1], **Annalina Mayer** [1], **Lei Zhang** [2], **David Stolz** [1], **Laura Müller-Pinzler** [1], **Sören Krach** [1]

[1] University of Lübeck, [2] University of Birmingham

Details

Objective: The formation and revision of self-beliefs heavily relies on integrating feedback into existing beliefs. Thus, self-beliefs are subject to constant (re)evaluation. Previous research has shown that belief formation processes are inherently biased, and once beliefs are firmly established, they tend to persist in the face of conflicting information. To gain a deeper understanding of these biases and the factors influencing whether a belief is revised or maintained, it is crucial to consider the role of initial expectations and confidence in these estimates. We hypothesize that as beliefs become established and confidence increases, belief updating shifts towards a more confirmatory integration of feedback.

Methods: In a within-subject design, 99 participants engaged in a trial-by-trial performance learning task. Initially, novel self-beliefs were established in a belief formation phase (session 1). These firmly established beliefs were then challenged with conflicting feedback in a belief revision phase (session 2). Computational modeling was employed to describe the initial formation and potential revision of self-beliefs. Learning biases from positive and negative feedback were computed and linked to expectations and confidence at different stages of learning.

Results: Our results indicate that feedback-based learning tends to be biased towards initial expectations, even when learning about relatively novel domains. Furthermore, experimentally induced self-beliefs were resistant to revision and this effect was stronger with increasing confidence in their expectation.

Conclusion: Essentially, self-beliefs appear to be formed and revised in a self-confirmatory fashion. This underscores the importance of considering individual initial beliefs when interpreting results on belief updating. Additionally, our results show the feasibility of iteratively establishing domain-specific and relatively novel self-beliefs under controlled conditions. This provides an opportunity to study belief updating and associated biases, motives, and affect that may arise during different phases of learning. Furthermore, it may prove beneficial for conducting effective research into interventions aimed at facilitating more effective belief revision in clinical, political, and economic contexts.

P1-H-50 - Credit assignment in learning under temporal uncertainty

Jiwon Park [1], **Dongil Chung** [1]

[1] Ulsan National Institute of Science and Technology

Details

Decisions made in the real world often result in temporally delayed outcomes. Due to such temporal uncertainty, making causal inferences when delayed outcomes are delivered becomes challenging. Despite its importance, it remains unclear how humans engage in learning from rewards delivered with uncertain delays and subsequently make decisions. Here, we used a probabilistic learning task where individuals (n = 60) were asked to make a series of choices between two options, the reward contingency of which could only be learned through experience. Importantly, the outcome of each choice was delivered after an uncertain delay sampled from a Poisson distribution. Regardless of the actual temporal uncertainty, individuals went through four different learning blocks, each with a temporal delay explicitly instructed to manipulate individuals' beliefs about the delay: zero trials (i.e., immediate reward delivery), one trial, two trials, or a random number of trials. Against alternative explanations, we found that individuals used both a model-based learning strategy, attributing credit for observed outcomes to their past choices based on their beliefs about the task's temporal structure, and a model-free learning strategy, attributing credit to their most recent choice when updating their expectations about future states ('future state values' in the temporal difference (TD) learning). Notably, we found that the extent to which individuals incorporated the future state values in the model-free learning strategy increased as their belief about the temporal delay became longer, and that this property was strongest when they were informed about the temporal uncertainty ($F(3, 214) = 7.89$, $p < 0.001$). Moreover, under informed temporal uncertainty, individuals who incorporated the future state values the most showed the highest performance in learning ($r = 0.46$, $p < 0.001$). The current study sheds light on the cognitive mechanisms underlying causal inferences under temporal uncertainty, providing insights into understanding maladaptive behaviors in decisions with long-term consequences, such as those related to health (e.g., obesity, substance abuse).

P1-H-51 - Paving the way for using neuroscience models in health economics: application of the active inference/free energy paradigm on health insurance choice

Eline Van Den Broek-Altenburg [1]

[1] University of Vermont

Details

Objective

There are increasingly more applications of choice modelling in health economics to better understand how patients value aspects of personal health, treatment options and different health insurance packages. There has been an increase in the adoption of ideas from behavioural economics and mathematical psychology. This has led to the investigation of behavioral concepts such as reference dependence, gains-loss asymmetry and zero price effects, and the development of new models such as regret minimisation. Economists are increasingly engaging with the field of neuroscience attempting to capture additional process data, mainly through neuroimaging techniques such as EEG or fMRI. The objective of this study is to analyze plan choice in Medicare in the U.S. using the Medicare Current Beneficiary Survey from 2007-2020 (data on beneficiary demographics, health, income and other attributes) with data on plan payment levels and measures of market competition.

Methods and Results

What is lacking thus far in health insurance choice is an engagement with the theories specifically developed in neuroscience, in relation to how the human brain perceives stimuli and how those stimuli are acted upon. One topic in particular has received extensive attention and generated much excitement in neuroscience: the free energy principle, developed by Karl Friston. The key notion behind the free energy principle is that the brain seeks to minimise surprise. We do this by decreasing the gap between our expectations of the world and our perceptions of the world, either by adapting our expectations or by taking actions to bring our surroundings (and hence perceptions) in line with our expectations. As a consequence of its top-bottom approach, the free energy principle can be used to explain multiple phenomena, including (but not limited to) memory, attention, value, reinforcement, salience, emotion and exploration behaviour. Typical choice models only include error terms to represent uncertainty from the viewpoint of the analyst. Our model looks at the incorporation of parameters for the uncertainty from the perspective of the decision-maker. Additionally, costs of deliberating (exploring) is taken into account through examination of the response times of a decisionmaker.

Conclusions

This study will be the first to apply preference accumulator models to both stated and revealed preference to analyze insurance choice. This work will form an important bridge between health economics and neuroeconomics.

P1-H-52 - Neural bases of Epiphany Learning: An fMRI Study Using a Race Game

Wei Chen [1]

[1] National Taiwan University

Details

This study investigates the neural basis of epiphany learning through an fMRI experiment using a race game. Epiphany learning, characterized by sudden comprehension or problem-solving, remains an underexplored area in cognitive neuroscience. Our experiment aimed to identify specific brain regions associated with this phenomenon.

Participants (N = 60) experienced several different race games designed to elicit epiphany moments. Through careful model predictions, we identified that 35% of the participants were epiphany learners. These individuals displayed distinct neural activation patterns compared to non-epiphany learners.

During moments of epiphany, significant activation was observed in the lateral frontal cortex and the left dorsolateral prefrontal cortex. The lateral frontal cortex is known for its role in complex cognitive processes, such as decision-making and problem-solving, suggesting its involvement in the cognitive restructuring that occurs during epiphany learning. Similarly, the left dorsolateral prefrontal cortex is associated with working memory and cognitive flexibility, supporting its role in adapting to new problem-solving strategies during epiphany moments.

Our findings highlight the importance of these brain regions in facilitating the cognitive shift that characterizes epiphany learning. The activation of the lateral frontal cortex and the left dorsolateral prefrontal cortex indicates a neural basis for reorganizing thought processes that lead to sudden problem-solving breakthroughs.

This study contributes to understanding the neural mechanisms underlying epiphany learning and provides a foundation for future research in cognitive neuroscience. By identifying specific brain regions associated with epiphany, we pave the way for exploring targeted interventions and training programs to enhance epiphany learning and problem-solving skills in educational and clinical settings.

Our research underscores the complexity of epiphany learning and the sophisticated neural processes involved, offering new perspectives on how the brain achieves sudden comprehension and innovative problem-solving.

P1-I-53 - Valuing, Fast and Slow

Eric Feddeck [1], **Inge Rabeling** [1], **Catalina Ratala** [1], **Alan Sanfey** [2]

[1] Radboud University, [2] Donders Institute

Details

Valuing, Fast and Slow

Eric Feddeck, Inge Rabeling, Catalina Ratala, and Alan Sanfey

Recent research in decision-making has supported an account of sequential sampling of information and accumulation of evidence until a decision is made. In this sequential process, it follows logically that not

all information about the value of choice options can enter the process at the same time but does so with different delays. Indeed, this has been investigated in a few experiments in the literature: Sullivan et al. (2015), Lim et al. (2018), Maier et al. (2020), and Fisher et al. (2023) have all found support for this account. The techniques used to investigate whether values enter the decision process at different times were twofold: In binary choice tasks, participants' mouse trajectories were analyzed as well as their reaction time and choice data computationally modelled.

1) The objective of this study (preregistered, *N* = 35) is to conceptually replicate the findings on differing processing speeds using a combination of separately established methods and analysis techniques used by different researchers. This will also provide an opportunity to validate the vastly different approaches taken to investigate this phenomenon.

2) In our paradigm, each of the 35 participants rated 70 food items on health and taste, before making 240 consequential choices between pairs of food items. Reaction time data, choice and mouse trajectory was collected. Mouse trajectories are analyzed by regressing the trajectory data on specific value attributes to determine the point at which they become significant predictors. Choices and reaction times are inspected by fitting a time varying drift diffusion model to the data and interpreting (among others) a parameter for the relative starting time of taste and health attributes before contributing to the evidence accumulation rate (drift rate) for each participant.

3) Results are expected soon.

4) The goal of using two different analyses of two streams of data (i.e., mouse trajectories and choice/RT) of behavior is to yield information on processing speeds of values and contribute to the body of literature on the sequential sampling account of value-based decision making. Contradictions in previous literature leave many questions unanswered as to the generalizability of evidence accumulation processes and this study provides the basis to tackle these questions. In addition to providing an opportunity to validate the vastly different approaches taken to investigate this phenomenon, this will act as a foundation for answering the broader research question of the project at hand: How is choice guided by the interaction of choice attributes and personal values?

P1-I-54 - Is the whole the sum of its parts? Neural correlates of consumer bundle valuation in humans

Logan Cross [1], **Ryan Webb** [2], **John O'Doherty** [3]

[1] Stanford University, [2] University of Toronto, [3] California Institute of Technology

Details

Humans have to evaluate options that contain multiple components, such as a meal, an investment fund, or a vacation package. From a normative standpoint, this can be accomplished via a hierarchical process of valuing the bundle's constituent items and then aggregating them to value the bundle: the

whole is the sum of its parts. This process is formalized in attribute integration theories of value, which propose that value is constructed by assigning values to its attributes and then aggregating them. Neuro-imaging studies have identified specialized cortical areas in lateral orbit-frontal cortex (lOFC) that encode stimulus attributes and exhibit functional connectivity to medial orbitofrontal cortex (mOFC) and ventromedial prefrontal cortex (vmPFC) for value integration (Suzuki et al., 2017; Lim et al., 2013). In this study, we investigated the fundamental question of how the value of bundles of items are represented in the brain. One possibility is that the attribute integration theory of value applies to valuing bundles of items, for which the components that need to be integrated simply represent a different level of abstraction: they are themselves composed of items. We might therefore expect to find that values of items and bundles exhibit a hierarchical representation over cortical area, perhaps continuing the previously-noted lateral to medial dissociation. Alternatively, the same regions may code for the value of items and bundles, suggesting a more generalized valuation process across different decision contexts.

Participants (N=14) underwent a deep-fMRI scanning protocol while we elicited valuations and choices for 40 individual items and bundles of those same items. Each trial involved a binary choice between an item or bundle and a reference monetary amount. Importantly, the single item and the bundled trials were randomly interleaved in the scanning session.

We find that subjects valuation behaviour systematically discounts bundles compared to the sum of individual item values. At the neural level, a distributed network of brain areas, including the vmPFC but also other parts of prefrontal cortex, compute the value of a bundle with the same value code employed for individual items. This suggests that these general value regions adapt to the item or bundle context rather then recruit other regions to separately code this hierarchy. Additionally, we tested various models of how this value code may adapt between levels of this hierarchy. Our findings suggest that the value representation undergoes a normalization process that actively re-scales the value code based on the distribution of values within the current context, as opposed to utilizing an absolute value code.

P1-I-55 - Acute stress impairs cognitive task performance at varying levels of computational hardness

Karlo Doroc [1], **Carsten Murawski** [1]

[1] University of Melbourne

Details

Objective: Acute stress triggers a cascade of physiological and psychological changes, including heightened perspiration, cortisol levels, and anxiety. Prior studies have focused on acute stress' effect on cognition in established paradigms such as executive functioning and economic uncertainty tasks. By comparison, there is little work on stress' effect in computationally hard tasks, even those which are ubiquitous and economically relevant. We posit that studying performance on computational problems is a more appropriate framework for those interested in understanding how a decision environment's complexity, or computational hardness, affects decision-making under acute stress. Here, we test for an effect of acute stress on human decision-making at varying levels of computational hardness in a canonical task well suited to economic contexts, the 0-1 Knapsack Decision Problem (KP), a constrained optimisation problem.

Methods: Participants (N = 42, mostly college students) completed two sessions in which they underwent either the Trier Social Stress Test (TSST; stress session) or the placebo-TSST (control session) prior to solving 72 trials of the KP. Trials were sampled using Franco et al.'s (2021) instance complexity measure, with half chosen to be high (low) in complexity. Behavioural measures included participant choice, time on task, and gaze tracking in the KP. Stress measures included salivary cortisol, skin conductance, negative and positive affect, and self-reported stress measures.

Results: We find that acute stress leads to impaired decision-making performance on the KP. Particularly among cortisol responders, higher cortisol levels were associated with lower accuracy ($\beta = -0.13$, $p < 0.01$) and higher time on task ($\beta = 0.07$, $p < 0.01$). Interestingly, we find no interaction effect between stress levels and task complexity on performance, which suggests that stress has a similar effect on decisions both high and low in complexity. Gaze tracking analysis revealed that attention was affected by stress, with stressed participants attending to more information, registering a shorter dwell time, and switching more frequently between different areas of interest; however, these differences did not explain the stress-induced differences in task performance.

Conclusion: Both acute stress and computationally hard tasks are ubiquitous in real-world decision situations, yet their joint effect on decision-making is not well understood. We are the first to show that acute stress impairs decision quality irrespective of the level of computational hardness, building a more nuanced understanding of how acute stress affects complex decision-making.

P1-I-56 - Neuro-metabolic correlates of the motivation to perform efforts

Nicolas Clairis [1], **Arthur Barakat** [1], **Carmen Sandi** [1]

[1] Behavioral Genetics Laboratory (LGC)

Details

People exhibit significant variability in their willingness to engage in both physical and mental efforts, yet the underlying neurobiological mechanisms driving this variability remain poorly understood. To address this gap, we investigated whether the level of certain metabolites in the dorsomedial prefrontal cortex/dorsal anterior cingulate cortex (dmPFC/dACC) and in the anterior insula (aIns), measured with proton magnetic resonance spectroscopy (^{1}H-MRS) in 75 individuals (39 females, aged between 25 and 40 years old) could predict inter-individual differences in effort-based decision-making during functional magnetic resonance imaging (fMRI). Employing computational modeling, machine learning and a mediation analysis, we found that baseline dmPFC/dACC metabolites levels were associated with individuals' propensity to engage in both physical and mental effort during fMRI. Specifically, a multivariate combination of glutamate, lactate and aspartate levels in the dmPFC/dACC predicted inter-individual differences in the drive to perform mental effort, while inter-individual differences in dmPFC/dACC lactate levels predicted variations in the willingness to engage in more intense physical effort. The latter was mediated by lactate impact on dmPFC/dACC neural activity during decision-making. Our findings underscore the significant influence of neuro-metabolism on motivated behavior.

P1-I-57 - Dynamic changes in Fatigue reduce the willingness to exert effort in Parkinson's Disease.

Selma Lugtmeijer [1], **Dariusz Pilucik** [2], **Annie Blake** [2], **Annika Kienast** [2], **Trevor Chong** [3], **Masud Husain** [2], **Matthew Apps** [1]

[1] University of Birmingham, [2] University of Oxford, [3] Turner Institute for Brain and Mental Health, Monash University

Details

Parkinson's disease (PD) is a progressive neurological disorder that is traditionally conceptualised as a condition affecting movement. Yet, compared to primary motor impairments, other symptoms including highly prevalent and debilitating levels of fatigue, are poorly understood. Previous work has suggested that PD patients suffer impaired motivation, that leads to a reduced willingness to exert effort for reward. However, this work has largely ignored the impact that different levels of accumulation of fatigue might have on motivation and thus effort-based choices in PD. In healthy adults, levels of fatigue arising from extended effortful behaviours impact cost-benefit valuation, with accumulated efforts increasing the discounting effect of effort on the value of rewards. Here, we use an experimental paradigm designed to uncover dynamic changes in the willingness to exert effort for reward in a group of Parkinson's disease (*N* = 37) patients off medication, and matched elderly controls (*N* = 39). On each trial they chose between resting for low reward (1 credit) or 'working', and exerting higher effort (30-48% of max grip strength) for higher reward (2 – 10 credits). Both groups showed effort discounting, but there was no difference between the two groups in effort-based choices in the first half of the experiment. Critically, only in later trials of the task, when fatigue would have accumulated, did PD patients show a reduced willingness to exert effort for reward. We conclude that dynamically changing levels of fatigue during extended effortful tasks may underlie reductions in the willingness to exert effort for reward in PD.

P1-I-59 - The explore/exploit trade-off drives behavioural variability in human and rat foraging decisions

Emma Scholey [1], **Matthew Apps** [1], **Mark Humphries** [2]

[1] University of Birmingham, [2] University of Nottingham

Details

Objective: Every day, animals make decisions about when to stop doing something rewarding, to travel to find rewards elsewhere. Such 'foraging' decisions, where the individual must choose between the resources being obtained in their current location (patch) or leaving and seeking better options elsewhere, are fundamental across species. Theoretical models from behavioural ecology predict that the optimal time to leave a depleting resource depends on the richness of the environment i.e. the possible long-run average reward rate. These models are deterministic, but it is well established that animals make stochastic choices. Here, we use computational modelling to understand how animals and

humans make stay-or-leave decisions in foraging-like paradigms, with the objective of understanding the drivers of stochastic choices during foraging.

Methods: We examined foraging decisions in three datasets across humans and rats (n = 39 humans [Le Heron et al. (2020)]; n= 29 humans [Contreras-Huerta et al. (2022)]; and n = 8 rats [Kane et al. (2019)]). In each experiment, subjects are presented with depleting rewards in patches, in environments with different average reward rates, and must choose when to leave. Through a combination of model simulation and formal model comparison, we tested a range of stochastic models to explain patch leaving decisions, as well as variability in an individuals' patch leaving.

Results: We show that the winning model was one that had separate parameters for a reward-based explore/exploit trade-off which differ between environments, as well as an exploration bias parameter that is unrelated to reward. Moreover, we found increased reward-based exploration in richer compared to poorer environments (Wilcoxon signed rank: Le Heron et al. (2020), $Z = -4.47$, $p < .001$, $r = 0.72$; Contreras-Huerta et al. (2022), $Z = -4.01$, $p < .001$, $r = 0.74$; Kane et al (2019), $Z = -2.52$, $p = .01$, $r = 0.89$) explaining why mean leaving times are on average shorter when foraging in rich environments. We also show that our model can account for counter-intuitive patterns of variability in subjects' leaving choices across the datasets. Critically, we show that variability in patch leaving is independent of the economic value of the patch.

Conclusions: We provide a model that robustly accounts for the explore/exploit trade off in decision-making, offering a computational account of foraging choices across species.

P1-I-60 - Efficient value coding emerges via synaptic plasticity in neural networks

Bo Shen [1], **Soomin Song** [1], **Robert Froemke** [1], **Kenway Louie** [2], **Paul Glimcher** [1]

[1] NYU Grossman School of Medicine, [2] New York University

Details

Objective: Humans and animals make decisions in a world full of redundant information. One of the options may carry information about the other. A neural circuit architecture of divisive normalization has been shown to eliminate redundant information across option values, and thus is thought to implement an efficient coding of decision information (Schwartz & Simoncelli, 2001; Louie et al., 2011; Glimcher & Tymula, 2023). Yet, it remains unclear how such architecture arises and whether it adapts to environments with varying degrees of redundancy. Here, we examine how the divisive normalization motif arises in neural networks with empirically derived, biological learning rules.

Methods: We quantified how the level of correlation across option values shapes network structure in a randomly connected spiking network of excitatory (E) and inhibitory (I) neurons governed by synaptic plasticity rules. Excitatory units implemented a classic Hebbian spike timing-dependent plasticity (STDP) rule, where synapses are potentiated after pre-post synaptic spike sequences and depressed after post-pre sequences. In contrast, inhibitory neurons implemented a recently discovered symmetric plasticity kernel, with synapses always potentiated regardless of the order of the pre and post spikes (Vogels et al. Science 2011; Field et al. Neuron 2020).

Results: In high redundancy environments, where two option values are correlated over time and therefore share a large amount of information, the network forms long-range E-to-I connections between the two option regions, establishing lateral inhibition that is mediated through local I-to-E connections – a typical motif of divisive normalization. In contrast, in low redundancy environments, where the two inputs are independent, the network exhibits reduced long-range E-to-I connections but maintains local I-to-E connections, favoring independent normalized coding of the two options.

Conclusions: These findings suggest that the divisive normalization motif can spontaneously arise when the neural network passively receives redundant value inputs; furthermore, the degree of divisive normalization adapts to the degree of information redundancy in the environment. More broadly, this work sheds light on the origins and plasticity of efficient value coding in neural circuits, highlighting the essential role of brain-environment interactions in shaping the decisional brain.

P1-I-61 - A Neurocomputational Basis of Effort-Based Decision Making in Substance Use

Kasey Spry [1], **Kenneth Kishida** [2], **Merideth Addicott** [2]

[1] Wake Forest University, [2] Wake Forest University School of Medicine

Details

Objective: Effort-based decision making (EBDM) evaluates the reward magnitude and probability in relation to the effort required to obtain it. If effort expenditure is perceived as increasingly costly, the reward becomes devalued. Dopamine modulates EBDM, as shown in human and animal studies, indicating that increasing or decreasing dopamine transmission enhances or diminishes (respectively) the willingness to expend effort for rewards. Dopamine transmission has also been shown to be affected by chronic substance use, such that over time production, transmission, absorption, and sensitivity to dopamine is altered. However, it remains unclear whether EBDM is affected by the chronic use of substances that affect dopamine. To address this knowledge gap, we investigated the neurocomputational basis of EBDM and its alteration among groups characterized by tobacco use and opioid use.

Methods: In this study, never-smokers (n = 19), ex-smokers (n = 23), cigarette smokers (n = 25), and opioid-dependent cigarette smokers (n = 30) completed an EBDM task that manipulated reward magnitude and probability for high and low effort task options. Computational models assessed trial-by-trial reward magnitude, reward probability, and effort expenditure's influence on the subjective value of the high and low effort options and task choice behavior. This quantitatively defined the participants' value of effort and the underlying influences of effort and reward to their choice behavior.

Results: Groups differed in the use of reward magnitude, reward probability, and effort expenditure to modify the subjective value of the choices. Never-smokers modified their subjective value more using effort expenditure (two-sample two-sided t-test; $p<0.001$), whereas ex-smokers, smokers, and opioid-dependent smokers modified their subjective value more using reward probability (two-sample two-sided t-test; $p<0.001$). Further, reward probability and reward magnitude were found to have a significant interaction for never-smokers (logistic regression model; $p<0.001$) and ex-smokers (logistic regression model; p-value = 0.004), indicating that the effect of reward probability on subjective value

depends on the reward magnitude in non-tobacco users, whereas this interaction was not present for smokers or opioid-dependent smokers.

Conclusions: Together, these results suggest that chronic tobacco use and opioid use affects EBDM. Understanding a neurocomputational basis of EBDM in chronic substance use may enhance our comprehension of both typical and atypical EBDM. These contributions hold the potential for translational utility in the diagnosis and treatment of substance use.

P1-I-62 - Gaze patterns associated with efficient few-shot rule-learning

Leah Bakst [1], **Caroline Ahn** [1], **Quan Do** [1], **Michael Pascale** [1], **Chantal Stern** [1], **Michael Hasselmo** [1], **Joseph Mcguire** [1]

[1] Boston University

Details

***Objective**:* Human problem-solvers are adept at using sparse information to infer abstract rules and regularities to guide future decisions. To better understand how humans extract relevant information and construct rule representations, we leveraged humans' propensity to explore the visual world during a few-shot rule-learning task. We identified gaze-based correlates of successful problem-solving, finding that more complex gaze patterns, which integrated evidence across multiple sources of information, were related to the efficiency of problem-solving.

***Method**:* Participants (n=32, 12 male, 20 female) learned arbitrary rules from three example input-output pairs, consisting of grids of colored tiles, and then applied each rule to a novel test input via an interactive interface. This task was inspired by the Abstraction and Reasoning Corpus (ARC); we adapted 40 rules from ARC for human participants to solve while undergoing eye tracking. Participants had up to two attempts and a 5-minute time limit to generate the correct solution for each problem, and were paid a bonus per correct solution. For our analyses, we decomposed the gaze trace into fixations and saccades. This allowed us to identify the region of interest (ROI) on the screen, if any, to which each fixation was directed, and categorize saccades as between-ROIs, within-ROIs, or neither. ROIs corresponded to the grid elements of the display: six grids from the three input-output example pairs, and two from the test input and output.

***Results**:* Participants performed well at the task (mean accuracy = 87%, range 53-100%), with better overall accuracy associated across participants with shorter average deliberation times (Spearman r = -0.427, p = 0.015), suggesting no speed-accuracy tradeoff. Deliberation time showed marked inter-individual variability even for the 8 problems that all participants correctly solved (range, 58.1 s). Across participants, shorter deliberation times were associated with more sophisticated gaze patterns, as measured by the proportion of saccades that were between-ROI and the Lempel-Ziv complexity of the sequence of fixations (between-ROI proportion: r = -0.474, p = 0.007; Lempel-Ziv complexity: r = -0.551, p = 0.001).

***Conclusion**:* Participants achieved a high level of success on a few-shot rule-learning task. There was substantial variability in response speed not accounted for by differences in accuracy, suggestive of

individual variation in problem-solving efficiency. Shorter deliberation times were associated with distinct gaze patterns, suggesting that sophisticated information-seeking strategies integrate evidence across examples in a more complex and efficient manner.

P1-I-63 - Neural dynamics of predicting others' choices

Erik Stuchlý [1], **Sebastian Gluth** [1]

[1] University of Hamburg

Details

Converging evidence suggests that people often predict the behaviour of other individuals by imagining what they themselves might do next in such a situation – essentially simulating the other's future hypothetical behaviour with their own mind. Similar conclusions have been drawn with respect to predicting the decisions of others, with multiple studies identifying the same cognitive and neural mechanisms being engaged during choice for oneself and when predicting the choices of others. In the current study, we sought to find additional evidence for this "simulation hypothesis" on the neural level during the decision process.

Participants (*N*=50; data collection ongoing) first made risky decisions for themselves on a varied set of options ("self stage"), then learned another agent's risk preference by observing their choices on a separate set of options, and finally predicted the choices of this agent on the set of options from the self stage ("prediction stage"). Furthermore, participants have learned and predicted for two agents – one with a risk preference similar to themselves, the other with a more dissimilar preference.

We hypothesized that, if the brain represents predicted decisions as its own, any time-locked ERPs which scale with decision difficulty during decisions for the self should also scale when predicting another individual's decisions, according to the difficulty as would be perceived by the other agent. To this end, we will apply a linear regression to the time-locked ERP data from the self stage, with trial-to-trial subjective value difference (computed with the participant's risk attitude parameter) as the predictor, to identify the channels and timepoints where the brain signal linearly scales with decision difficulty. Subsequently, we will fit another regression model to the prediction stage data from these channels and timepoints, with the predictor of subjective value difference (computed with the risk attitude parameter used to make predictions for the other agent). Additionally, and in line with some previous literature, we also expect to find early signals distinguishing between whether the choice is being made for oneself, a similar other or a dissimilar other, particularly around the temporo-parietal region. We will test this hypothesis by comparing time-locked ERPs from the self stage and the two prediction stages.

The poster will communicate results of these analyses, with the aim to increase the understanding of the neural and cognitive dynamics underlying the prediction of others' decisions and assess the extent to which they overlap with standard decisions for oneself.

P1-I-64 - A neurocomputational framework for multi-attribute value construction

Raphael Le Bouc [1], **Gilles De Hollander** [1], **Marcus Grueschow** [1], **Christian Ruff** [1]

[1] University of Zürich

Details

Making good decisions often involves comparing options that differ on multiple dimensions, such as when choosing the best house or the most advantageous job. These decisions are believed to rely on neural signals in the ventromedial prefrontal cortex (vmPFC) that integrate the value of the different attributes of the options. Recent research suggests that these value signals are computed as a precision-weighted sum of the attribute values, with uncertain attributes contributing less than certain ones. However, it is still unclear how this uncertainty is derived from our senses, how it is represented in the brain, and how it influences value estimation.

We propose that value is constructed through a hierarchical Bayesian inference process that combines noisy sensory evidence about attribute values. We hypothesize that the vmPFC and sensory cortices employ similar probabilistic population coding principles, encoding value and percepts, respectively, along with their uncertainty as probability distributions in neuronal population activity.

We tested this hypothesis with fMRI (N=40) during a value rating task involving multisensory stimuli with two attributes: the orientation of moving dots and the auditory frequency of pure tone sequences. Both attributes were mapped to attribute values that jointly determined the payoff resulting from the stimulus. By manipulating the precision of the visual attribute through changes in motion coherence, we demonstrated that value inference reflects the weighting of sensory evidence proportional to its precision. This was evidenced by a significant interaction between the value of the visual attribute and its sensory precision on value estimation (multiple logistic regression, $\beta = 2.12 \pm 0.72$, $p < 0.001$).

Moreover, using a population receptive field model and a Bayesian decoder (Berretto-Garcia et al., 2024; van Bergen et al., 2015), we found that perceptual uncertainty decoded in the visual cortex explained how participants weighted the visual attribute during value inference (multiple logistic regression, $\beta_{\text{MixedEffects}} = -0.017 \pm 0.007$, $p < 0.014$). We further investigated whether probabilistic neural value signals decoded in the vmPFC reflect precision-weighted integration of sensory evidence about value.

Our findings provide insight into how noisy sensory information about multiple attributes is integrated into value signals, shed light on the Bayesian computational principles by which values are inferred from multiple sources of information, and establish a link between these computations and neural population coding principles.

P1-I-65 - Avoiding physical effort may depend on the motivation to exert cognitive effort into difficult choices

Jamie Talbot [1], **Anna Krnic** [1], **Matthew Apps** [1]

[1] University of Birmingham

Details

People are averse to exerting both cognitive and physical effort. In decision-making paradigms people choose to avoid higher effort unless it can obtain larger rewards, an aversion which is greater in people higher in apathy – defined as a reduction in goal-directed behaviour. Yet, making decisions can be difficult and thus may also be cognitively effortful. Therefore we ask - are people motivated to exert cognitive effort to avoid physical effort? Here we use a novel paradigm and drift-diffusion modelling to assess how people modulate cognitive control when choosing between multiple physically effortful options and test its association with different domains of apathy.

Participants (n=68) were presented with a task in which they had to choose between stimuli representing physical effort, as well as completing apathy questionnaires. The "value" of these options was one of seven different levels of physical effort (20-70% of max grip strength) that were learned to a pre-set criteria of accuracy during a training session. Unlike previous decision tasks, the reward for picking each option was the same, thus participants were incentivised to select the lowest effort option. On each trial they were presented with either two or three options, that were varied in how close together in value the best two options were. In addition, on a subset of trials they were required to rate (between 0-10) how effortful they found the choice. Thus, the difficulty of decisions was manipulated both in terms of value difference and number of options, and we could assess how these affected both reaction times and ratings of effort.

We show that participants are slower ($p<0.001$) and report higher effort required when the value difference of the two best options is smaller ($p=0.02$). They were slower ($p<0.01$) in three versus two option choices, although did not report these as more cognitively effortful ($p=0.7$). In addition, we found that the cognitive/executive apathy domain was associated with slowed response times ($p=008$), lower accuracy ($p=0.05$), greater indecision (higher rate of choice failure; $p=0.008$) and lower response thresholds when modelled in a DDM. The same pattern was not shown for the action initiation domain of apathy.

These results demonstrate that people find choosing between physically effortful behaviours cognitively effortful, especially when those effort levels are close together and when there are multiple effortful behaviours offered. This suggests that the motivation to exert cognitive effort may be key for understanding how people decide to pursue physically effortful behaviours, and that this motivation may be reduced in people with cognitive apathy.

P1-I-66 - Specific Sensitivity to Rare and Extreme Events: Quasi-Complete Black Swan Avoidance vs Partial Jackpot Seeking in Rat Decision-Making

Mickael Degoulet [1], **Louis-Matis Willem** [2], **Christelle Baunez** [1], **Stephane Luchini** [3], **Patrick Pintus** [4]

[1] Institut Neurosciences Timone, [2] Centrale Meditérannée, [3] Aix-Marseille University, [4] Centre national de la recherche scientifique (CNRS)

Details

(1) study's objective: Most studies assessing animal behavior under risk rely on probabilities that are typically larger than 10%. To study Decision-Making in uncertain conditions, we explore a novel experimental and modelling approach that aims at measuring the extent to which rats are sensitive - and how they respond - to outcomes that are both rare (probabilities smaller than 1%) and extreme in their consequences (deviations larger than 10 times the standard error).

(2) methods used: In a four-armed bandit task, stochastic gains (sugar pellets) and losses (time-out punishments) are such that extremely large - but rare - outcomes materialize or not depending on the chosen options.

(3) results: All rats feature both limited diversification, mixing two options out of four, and sensitivity to rare and extreme outcomes despite their infrequent occurrence, by combining options with avoidance of extreme losses (Black Swans) and exposure to extreme gains (Jackpots). Notably, this sensitivity turns out to be onesided
for the main phenotype in our sample: it features a quasi-complete avoidance of Black Swans, so as to escape extreme losses almost completely, which contrasts with an exposure to Jackpots that is partial only. The flip side of observed choices is that they entail smaller gains and larger losses in the frequent domain compared to alternatives. We have introduced sensitivity to Black Swans and Jackpots in a new class of augmented Reinforcement Learning models and we have estimated their parameters using observed choices and outcomes for each rat. Adding such specific sensitivity results in a good fit of the selected model - and simulated behaviors that are close - to behavioral observations, whereas a standard Q-Learning model without sensitivity is rejected for almost all rats. This model reproducing the main phenotype suggests that frequent outcomes are treated separately from rare and extreme ones through different weights in Decision-Making.

P1-I-67, PS.01.01 - Sex differences in change-of-mind neuroeconomic decision-making is modulated by LINC00473 in medial prefrontal cortex

Romain Durand-De Cuttoli [1], **Orna Issler** [2], **Benjamin Yakubov** [3], **Nusrat Jahan** [3], **Aisha Abid** [3], **Susanna Kasparov** [3], **Kastalia Granizo** [3], **Sanjana Ahmed** [3], **Scott Russo** [1], **Eric Nestler** [1], **Brian Sweis** [1]

[1] Icahn School of Medicine at Mount Sinai, [2] New York University, [3] Hunter College

Details

Ruminating on the past when deciding to change one’s mind is a complex cognitive phenomenon involving a continuous re-appraisal of the trade-off between past costs and future value. Recent work modeling this behavior across species established contributions of alterations in this decision-making process to psychopathology. The incidence of depression is two times higher in women than in men and can manifest different symptomatology, including negative rumination. Strides made in neuroeconomics have segregated stages of a decision stream into its component parts, isolating properties of change-of-mind re-evaluative decisions, and pointing to the medial prefrontal cortex (mPFC) as a critical hub for regulating reward value, self-control, and mood. We recently demonstrated that mPFC activity was

causally linked to change-of-mind decisions. Recent transcriptomic analyses of human postmortem mPFC tissue obtained from individuals diagnosed with depression revealed a female-specific decrease in the long intergenic non-coding RNA: LINC00473 expression in the mPFC. We investigated in mice the actions of mPFC LINC00473, whose role in cognition has not yet been studied. While LINC00473 is not normally expressed in rodents – it arose in the primate lineage – expressing LINC00473 in mouse mPFC neurons demonstrated pro-resilient effects in females, but not males, using simple stress assays. Here, we characterized complex decision-making behavior in male and female mice longitudinally in our neuroeconomic foraging paradigm, Restaurant Row, following virus-mediated LINC00473 expression in mPFC neurons. On this task, mice foraged for their primary source of food among varying costs (delays) and subjective value (flavors) while on a limited time-budget during which decisions to accept and wait for rewards were separated into discrete stages of primary commitments and secondary re-evaluations. We discovered important sex differences in decision-making behavior. LINC00473 expression selectively influenced multiple features of re-evaluative choices, without affecting primary decisions, in female mice only. These behavioral effects included changing how mice (i) cached the value of the passage of time and (ii) weighed their history of economically disadvantageous choices. Both processes were uniquely linked to change-of-mind decisions and underlie the computational bases of distinct aspects of regret-related counterfactual thinking, potentially restoring neuroeconomic vulnerabilities in females. These findings reveal a key bridge between a molecular driver of stress resilience and psychological mechanisms underlying sex-specific decision-making proclivities.

P1-J-68 - How hard (physical) effort feels is dependent on the history of exertion and rests

Tanja Müller [1], **Joseph Milton** [2], **Masud Husain** [2], **Matthew Apps** [3]

[1] University of Zürich, [2] University of Oxford, [3] University of Birmingham

Details

The amount of effort incurred by a task plays a pivotal role for our decisions as to whether we engage in it. Most work on effort-based decisions assumes that effort is costly, with the cost static over time. In contrast, neuro-cognitive accounts propose that fatigue and effort are closely linked, yet the computational basis of this coupling is unclear. We recently developed a computational account of perceived fatigue, suggesting that it fluctuates moment-to-moment as a function of the history of exertion and rest, and showed that it can predict trial-by-trial effort-based decisions. However, whether subjective perceptions of effort, rather than decisions, fluctuate in a similar way remains unknown. The aim of the present work was to examine the extent to which perceived fatigue and effort are underpinned by the history of exertion and rest and to test our computational framework.

Three experiments were developed, investigating how perceived effort and fatigue fluctuate over the course of a task as a function of effort exerted, rest taken, and reward obtained. In each experiment, participants performed a physically fatiguing task where they exerted different levels of effort (%ages of their maximum grip force) or briefly rested, for different amounts of reward (number of credits awarded for successful exertion). Crucially, participants also gave trial-by-trial self-report ratings of perceived effort (Study 1, *n*=41) or fatigue (Study 2, *n*=40), with a final study controlling for any effects of reward incentivisation on force and subsequent ratings.

Mixed effects models revealed that the perception of effort not only depended on the effort level ($\chi^2(1)=4191$, $p<.001$) but that it also increased over repeated exertions ($\chi^2(1)=241$, $p<.001$). Computational modelling showed that perceived fatigue as well as perceived effort were susceptible to recoverable and unrecoverable components as implemented in our hypothesised model. Whilst higher reward in Study 1 and 2 lead to slightly higher perceived effort ($\chi^2(1)=22$, $p<.001$) and fatigue ($\chi^2(1)=12$, $p<.001$), overall, such effects seemed to be due to participants overexerting force when high rewards were at stake and to the way in which reward information was presented.

These experiments demonstrate a close link between the effects of effortful exertion and brief rests on perceived effort and fatigue and highlight that an altered sensitivity to effort costs may be a crucial feature of momentary fatigue fluctuations. In addition, the findings have implications for the design and interpretation of studies on effort-based decision making, information seeking and learning, emphasising the dynamic nature of effort costs.

P1-K-69 - Improving models of attention: Distinguishing between agent-specific and decision-specific channels of attention in risky decision-making

Jan Engelmann [1], **Alejandro Hirmas** [2], **Joel Van Der Weele** [1]

[1] University of Amsterdam, [2] Universiteit van Amsterdam

Details

Economists have become increasingly interested in using attention to explain behavioral patterns both on the micro and macro level. This has resulted in several disparate theoretical approaches. Some, like rational inattention, adopt a “goal-directed” model of attention, where what is attended to is selected by the decision maker. Others, like salience theory, assume that attention is “stimulus-driven” and influenced largely by contextual factors. This distinction is fundamental for the economic implications of attention, but so far there is little understanding of their relative importance. In recent eye-tracking studies we distinguish between the impact of agent-specific attention and decision-specific variation in attention, which parallel the distinction between goal-directed and stimulus-driven attention that is commonly made in cognitive science. We verify our framework in a number of experimental contexts: (1) an eye-tracking experiment on risky choice; (2) a follow-up experiment that determined the causal influences of attention on risky choice; (3) a recent application in patients with gambling disorder (GD), and (4) an extension in a multi-attribute context involving sustainable decisions. Jointly, our results show significant improvements in model fits, underlining the importance of differentiating between agent-specific attention and decision-specific variation in attention in identifying the underlying cognitive mechanisms involved in economic choice. We expect future eye-tracking research to benefit from adopting this dual-channel approach to attention particularly because of its potential to enhance the precision of statistical modeling.

Poster Session II

P2-A-70 - Attentional mechanisms on gender homophily bias in equity crowdfunding: an eye-tracking study

Marco Barone [1], **Candida Bussoli** [1], **Lucrezia Fattobene** [2], **Aiqing Ling** [3]

[1] LUM Giuseppe Degennaro University, [2] University of Rome Tor Vergata, [3] University College Dublin

Details

Motivation. Equity crowdfunding (EC) has grown remarkably as an alternative investment form compared to traditional financial markets. Research on the success factors of EC campaigns has primarily focused on project creators and platform cues, largely overlooking investors' characteristics. Previous studies that looked into this last aspect have identified the gender homophily bias: female investors are more likely to invest in female (co)founder-dominated projects. However, the mechanism underlying this gender homophily bias remains to be elucidated. This study investigates the gender homophily bias by analysing investors' attentional mechanisms towards team composition and its influences on their EC decisions.

Methods. The experiment employed a 3 (EC team gender composition: all-male, all-female, balanced) × 2 (sustainability-oriented vs technology-oriented category) within-subjects design. Experimental stimuli comprised real EC campaign information collected from a large online platform composed of a total of 24 EC projects. During the experiment, the project information was randomly shown to participants for a maximum of 60 seconds, followed by a question regarding project attractiveness (att) on a 7-point Likert scale. 139 US participants (59 males, M_{age} = 54.79) joined in this experiment. We applied online eye-tracking to measure their visual processing of project information. The analysis employed ordered logit models.

Results. Controlling for age, education, income, financial literacy, and financial self-efficacy, female investors assessed EC campaigns more positively (M_{att} = 4.66) than male investors (M_{att} = 4.42, p < .00). This effect is mediated by the number of revisits towards information displayed in the EC project. We confirmed the gender homophily bias: female investors compared to men are strongly influenced by the presence of campaigns with all female teams, increasing the likelihood of assigning high scores compared to balanced (OR_{att} = 1.27, p < .00) and all-male (OR_{att} = 1.23, p = .01) compositions. Eye-tracking data indicates that this effect for all-female projects is mediated by the dwell times towards team composition.

We also found that gender homophily bias is influenced by campaign category. Although female investors prefer sustainability campaigns to technology ones (M_{sust} = 4.92, M_{tech} = 4.39, p < .00), they exhibit a stronger homophily bias when evaluating technology projects: polarized gender compositions receive higher ratings ($OR_{all\text{-}fem}$ = 1.44, p < .00; $OR_{all\text{-}male}$ = 1.21, p = .07) compared to gender diversified teams. For sustainability campaigns, the homophily bias is not significant ($OR_{all\text{-}fem}$= 1.15, p = .21).

P2-A-71 - Socioeconomic Status and Risk Taking Behaviors: Evidence from Online Gambling

Sally Gainsbury [1], **Hyundam Je** [1], **Emily Shaw** [1], **Agnieszka Tymula** [1]

[1] University of Sydney

Details

Objective: Numerous previous studies have suggested that lower socioeconomic status (SES) is associated with poor decisions, potentially perpetuating a cycle of poverty. In this project, we explore the relationship between SES and anomalies in risk-taking behavior. Specifically, we examine whether different SES levels are linked with well-known anomalies in risk-taking behaviors, such as the house money effect and loss chasing.

Methods: We utilize individual-level panel data from 10,000 individuals provided by a prominent Australian online sports-betting provider. This data includes details about betting decisions (e.g., odds, amount staked, outcomes), transactions (e.g., deposits, withdrawals) , and postcodes. To measure SES, we match each individual's postcode with Australian Bureau of Statistics (ABS) census data on neighborhood disadvantage. Using this data, we investigate whether there are anomalies in risk-taking behaviors and whether these are associated with SES.

Results: We find that lower SES individuals bet more and lose more, compared to higher SES individuals. Furthermore, anomalies in risk-taking behavior, such as the house money effect and loss chasing are more pronounced in lower SES individuals as well.

P2-B-73 - The Influence of Regret on Insurance Purchase Decisions: Insights from an ERP Study

Xiaolan Yang [1], **Mengru Ji** [1], **Mei Gao** [1]

[1] Shanghai International Studies University

Details

Insurance purchase is a prevalent financial decision that provides individuals with protection against unforeseen risks and helps mitigate potential financial losses in various aspects of life. Recent studies emphasize that regret plays a crucial role in insurance purchase decisions of individuals, based on anticipated feelings of remorse or dissatisfaction with their decisions. In the realm of insurance, there are two typical types of regret: "Under-Commitment Regret" (which measures the regret from no insurance coverage to offset incurred losses, leaving financially vulnerable after experiencing damages), and "Over-Commitment Regret" (which measures the regret from having purchased insurance without experiencing any damages, perceiving it as unnecessary and leading to a perceived wasteful expenditure). Our study aims to directly distinguish the intensity between these two regrets using event-related potentials (ERPs) within the domain of insurance purchasing decisions. We also investigate how regret emotions impact subsequent insurance purchase decisions. We employ an EEG experiment comprising 180 trials of insurance purchasing decisions. In each trial, participants choose to buy an insurance or not and subsequently receive feedback on whether losses occur. They should assess how regret they feel and then proceed to make insurance purchase decisions in the subsequent trial. Behavioral results show that "Under-Commitment Regret" is significantly stronger than "Over-Commitment Regret" according to self-reported regret. The difference in self-reported feelings of regret between occurrences of damage and no damage in no insurance purchase condition is significantly stronger than the difference observed between occurrences of no damage and damage in insurance purchase condition. There is a significantly positive relation between current feelings of regret and subsequent insurance purchases.The stronger the feeling of regret experienced after receiving feedback, the greater the likelihood of subsequent decisions to purchase insurance. Consistent with behavioral

results, we observed that the d-FRN (feedback-related negativity difference) between occurrences of damage and no damage in no insurance purchase condition is significantly more negative than the d-FRN between occurrences of no damage and damage in insurance purchase condition, indicating that "Under-Commitment Regret" is stronger than "Over-Commitment Regret". Armed with our findings, individual can select insurance products that offer the right level of coverage without over-committing, leading to a more satisfying and regret-free insurance experience.

P2-B-74 - Low level visual and auditory features predict preference for commercials

Inbal Gur Arie [1], **Dino Levy** [1], **Valentina Parada** [1]

[1] Tel Aviv University

Details

In recent years the use of videos for marketing and other domains has become extremely prominent and with it the budget for video advertisements. This has increased the need to investigate what attributes and features makes a video more likable and therefore successful. Previous work has shown that low level features are related to the saliency of the stimuli (Einhäuser et al., 2024), influence the preference of simple visual stimuli (McManus et al., 1981) and the preference for art and photographs (Iigaya et al., 2021). However, it is still not known if low level features influence the preferences of videos in general and of commercials in particular. In this study, we examined if low level visual and auditory features, extracted from video commercials for consumer products (such as colorfulness, length, scene frequency, zero crossing rate, spectral bandwidth, etc.), are related to the overall preference of those commercials. First, subjects (n=87) watched 100 video commercials in a randomized order. After watching each commercial, they stated their willingness to pay (WTP) for the product appearing in the commercial and indicated how much they liked the commercial. We extracted the low level visual features using a video mining tool developed by Schwenzow et al., (2020) and extracted the low level auditory features using the Librosa library in Python. To examine if low level visual and auditory features predicted the overall ranking of commercials we performed a Leave One Commercial Out Ridge regression. We found that both low level visual (R^2=0.28) and auditory (R^2=0.09) features can predict the likability rankings of the commercials and the average WTP for the products appearing in them. Moreover, combining both visual and auditory features increased the prediction success (R^2=0.31) suggesting that these two domains have unique contributions to the prediction. We have identified several important features that were strongly related to commercials' rankings and WTP of the products. Specifically, these features were scene frequency for the visual features and the variance of zero crossing rate for the auditory features. These results reveal that low level visual and auditory features influence the likability of commercials. These findings could potentially guide marketers and creative teams during the development process of novel commercials.

P2-B-75 - Source Memory Is More Accurate for Opinions Than for Facts

Daniel Mirny [1], **Stephen Spiller** [2]

[1] IESE Business School, [2] University of California, Los Angeles

Details

Effective communication relies on consumers remembering, sharing, and applying relevant information. Source memory, the ability to link a claim to its original source, is an essential aspect of accurate recall, attitude formation, and decision making. We propose that claim objectivity, whether a claim is a fact or an opinion, affects memory for the claim's source.

This proposal follows a two-step process: (i) Correctly recalling the original source of a given claim relies on the strength of the link formed between the source and the claim during the initial encoding of information (Greene et al. 2021; Mitchell and MacPherson 2017; Pham and Johar 1997); (ii) Opinions provide more information about a source than do facts (Heiphetz et al. 2014).

Across thirteen pre-registered experiments (*N*=7,510) and a variety of consumer domains, we investigate the effect of claim objectivity on source memory. The design used in each experiment was drawn from the source memory literature (e.g., Kassam et al. 2009). Stimuli were pre-tested to ensure that differences in claim objectivity were not confounded by differences in claim emotionality, valence, or arousal.

In six experiments, we find that source memory is more accurate for opinions than for facts, with no consistent effect on claim recognition memory (Exp 1 (*N*=399): *b*=12.47, *t*(321)=10.54; Exp 2a (*N*=501): *b*=2.63, *t*(412)=2.23; Exp 2b (*N*=504): *b*=3.79, *t*(440)=3.43; Exp 2c (*N*=503): *b*=4.08, *t*(435)=3.40; Exp 2d (*N*=501): *b*=3.51, *t*(464)=3.40; Exp 3 (*N*=606): *b*=5.33, *t*(560)=3.68).

We find support for the proposed process by manipulating facts to be more informative about sources, increasing source memory accuracy (Exp 4 (*N*=403): *b*=3.99, *t*(305)=3.49) and opinions to be less informative about sources, decreasing source memory accuracy (Exp 5 (*N*=1,213): *b*=1.00, *t*(1076)=1.05).

When forming inferences and seeking advice from sources, participants integrate new evidence more accurately based on sources' previously shared opinions than facts (Exp 6a (*N*=640): *b*=4.07, *t*(590)=4.00; Exp 6b (*N*=639): *b*=2.78, *t*(589)=2.61).

In three supplemental experiments, substantial changes to the experimental design resulted in attenuations of the main effect when using complex stimuli (Exp S1 (*N*=499): *b*=-.67, *t*(321)=-.96), when testing source memory using a cued rather than a full recall task (Exp S2 (*N*=501): *b*=.03, *t*(401)=.03), and when sources were periodical publications rather than humans (Exp S3 (*N*=601): *b*=-.89, *t*(551)=-.85).

Our results indicate that opinions are more likely to be accurately attributed to original sources than are facts.

P2-B-76 - Predicting Consumer Purchase Intention with Different Payment Methods: A Neuroforecasting Approach with Regression and Machine Learning

Aiqing Ling [1], **Manlin Wang** [2]

[1] University College Dublin, [2] Hangzhou Vocational and Technical College

Details

Motivation. Consumer neuroscience have shown that neurophysiological data, measured via fMRI and EEG, can forecast market-level consumer behaviours. Prior neuroforecasting research has typically employed one class of prediction technique (e.g., regression, machine learning or deep learning). Less is known if different analytic methods may have varying forecasting accuracy based on a common dataset. In addition, it is arbitrary in prior research to select different numbers and locations of EEG electrodes when conducting forecasting analysis. In other words, the spatial sensitivity of EEG inputs on forecasting remains unexplored. Finally, while fMRI-based neuroforecasting research has pinpointed two critical brain regions that are pivotal to forecasting, EEG-based studies have reported *diverse* components that are linked to different aggregate-level outcomes. It is unclear what the contributing weights of these EEG components are in a horse racing forecasting competition, assuming they all possess some prediction power. This study is motivated to address these three limitations in the literature.

Methods. We draw on published data (N=66, 160 trials per participant) to test how EEG signals evoked when knowing the payment options could forecast participants' subsequent purchase intention of high- and low-priced products. We incorporated theory-implied ERPs (P200, N270 and LPP) together with five power bands (alpha, beta, theta, gamma, delta) using a set of regression and machine learning techniques (i.e., logistic regression (LR), K-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), random forest (RF), and Support Vector Machines (SVM)) in the analyses.

Results. Using a 10-fold cross-validation approach for training, when all above EEG features from 60 channels were considered, all models forecasted significantly better than chance (*ps* < .001). However, machine learning models significantly outperformed a simpler LR (*ps* < .001). Among the machine learning models, RF, and SVM are better than LDA and KNN. When reducing EEG input channels from 60 to 32, 16 and 4, forecasting accuracies for all machine learning models were penalized (*ps* < .001) but not for LR. This may indicate that machine learning methods are sensitive to the reduction of EEG inputs and require a good amount of data to generate better forecasting. We evaluated how much different EEG features contributed to forecasting accuracy based on the random forest model with 16 channel inputs. We found that two power bands (gamma and theta) and the combination of two ERP components (LPP and N270) have highest weights on forecasting accuracy.

P2-B-77 - Using fMRI neural responses to distinguish AI-generated and real images

Thomas Henning [1], **Antonio Rangel** [1], **Colin Camerer** [1]

[1] California Institute of Technology

Details

AI image generating algorithms have taken the world by storm, with AI images being used as substitutes in movies, advertisements, and even dating profiles. As these image-generating algorithms have improved, many humans find the two classes of images to be indistinguishable. This has raised concerns for an upcoming crisis in social disinformation, including the phenomenon of “fake news.” In response, there is a need to design technologies that can help distinguish AI-generated and real images. In this study we investigate the role of neural responses in developing such technologies.

Previous neuroforecasting studies have shown that neuronal data can outperform human behavior in predicting real-world outcomes in certain domains (Knutson & Genevsky 2018). Here we build on this methodology to test if fMRI BOLD responses to face images can help distinguish AI-generated vs. real images above an individual's behavioral judgments.

The study was carried out in three stages:

In stage 1, we collected 300 real face images, and used AI to generate 300 face images. Each AI image was generated using a prompt inputted into the Midjourney v 5.1 algorithm. We ran an online experiment to measure the “image difficulty”, the accuracy of human participants in identifying the type of image. We then used this data to create our stimulus set for the fMRI experiment, which consisted of 100 “hard” images that were classified correctly at roughly chance (50/50), and 100 “easy” images that were classified correctly 70% of the time. Each class contained an equal number of real and AI-generated images.

In stage 2, we carried out an exploratory study (N=25) in which participants performed a classification task to distinguish AI vs Real faces. Each face was shown in the scanner for up to 3s and subjects had to predict its type (real vs. AI-generated). Task performance was incentivized with a monetary bonus for correct performance.

As a preliminary analysis, we fit a GLM of the BOLD responses to real and AI-generated images and found increased response for Real over AI images in the FFA and the vPFC regions, and increased activity for AI over Real images in the Retrosplenial, Precuneus, Anterior Cingulate, IFG, Ventral Tegmental regions.

We are also fitting and comparing machine-learning models that use both the behavioral and neural responses to predict the image type at the single image level, both at the individual and the group levels. Preliminary results are promising and will be reported in more detail at the conference.

Finally, in stage 3 we will use a new dataset of the same task to test out-of-sample all of the exploratory analyses and methods developed in stage 2, which we view as a critical step in developing reliable neurotechnologies.

P2-B-78 - A novel tool for measuring the continuous changes in value

Inbal Gur Arie [1], **Dino Levy** [1]

[1] Tel Aviv University

Details

Value is dynamic, shifting with internal state or external contexts. Historically, decision-making research has mostly focused on fixed and discrete stimuli in choice scenarios. Even with dynamic stimuli like videos, valuation and choice are mostly assessed discreetly (actual choices, WTP, ratings, etc.). However, the value attributed to a dynamic stimulus likely changes as a function of time. Such moment-by-moment valuation changes are not captured by discrete assessments typically made at the end of the stimulus. Our study developed a new tool that measures value continuously on a moment-by-moment basis using video commercials as the dynamic stimuli and systematically measures how value changes over time.

To build this tool, we presented subjects with 10 video commercials ranging from 30 to 90 seconds. During viewing, subjects continuously indicated their liking on a scale from 0 to 100. After each video, subjects answered two questions in a counterbalanced order: their overall liking on a scale of 0 to 100, and their willingness to pay (WTP) for the product featured in the commercial on a scale of 0 to 100 tokens. Each video was shown 3 times in randomized order.

We checked the tool's reliability by examining subjects' internal consistency, calculating pair-wise correlations between the continuous liking scores across the three repeats of each video. Subjects showed high consistency in their liking ratings ($t(803) = 44.12$, $p < 0.0001$, Cohen's $d = 1.55$). We further checked the tool's validity by correlating the overall liking score given at the end of viewing with the average continuous liking for each video, finding an average correlation of 0.86 across subjects. A significant mixed models linear regression corroborated the effect ($p < 0.001$). WTP and liking were only moderately related (average correlation across subjects was 0.47). Furthermore, there is no agreement between subjects in their dynamic likings of the commercials (ICC = 0.01, $p = 0.85$) and not in their averaged likings (ICC = -0.01, $p = 0.9$), demonstrating that our tool captured the variability across subjects in preferences towards the commercials. However, several time points within the commercials showed high agreement in liking scores across subjects, indicating the tool's ability to identify key factors within the commercials.

To conclude, we successfully built a novel tool with high validity and reliability that continuously measures value as subjects view dynamic stimuli. This tool enables the continuous tracking of the neural correlates of value changes and the investigation of the most engaging or dominant parts of the stimuli, characterizing them accordingly.

P2-C-79 - Humans decide to help others more often in poor environments

Todd Vogel [1], **Jo Cutler** [1], **Tabitha Hogg** [1], **Nima Khalighinejad** [2], **Luke Priestley** [2], **Neil Garrett** [3], **Matthew Apps** [1], **Matthew Rushworth** [2], **Patricia Lockwood** [1]

[1] University of Birmingham, [2] University of Oxford, [3] University of East Anglia

Details

Prosocial behaviours—decisions that help others—are essential for solving global challenges and for a cohesive society. Researchers have probed these using economic games or tasks where people decide between helping or not. However, in everyday life, current behaviours are interrupted with alternative options. Theories from ecology suggest that organisms are more likely to interrupt their behaviour to act when they are in poor compared to rich environments, and contested findings from social psychology suggest that poorer people are more prosocial.

Here, we measured how willing people are to interrupt ongoing behaviours to help others or themselves and experimentally manipulated environmental richness. People (study 1 (online), *n*=237 pre-registered, study 2 (online) pre-registered replication, *n*=219; study 3 (in-person), *n*=54) watched a movie whilst encountering opportunities to interrupt it to benefit another person, or on separate trials, themselves. Crucially, participants decided in two environments. In poor environments, average reward values and probabilities of opportunities were lower. In rich environments, average reward values and probabilities of opportunities were higher. Across all studies we demonstrate a robust ecological influence on decisions that benefit others. People were more willing to interrupt their behaviour to help others in poor compared to rich environments, and this influence was stronger compared to decisions that benefitted oneself (*p*s < .027).

Computational modelling of choice behaviour revealed that the opportunity costs of different environments were encoded distinctly for others and self (*p*s < 0.001). Value sensitivity when deciding for others in poor environments reached the same level as decisions to help oneself in rich environments (BF_{01} = 18.98). Across individuals, distinct factors of utilitarianism (*p*s < 0.007), and empathy/motivation (*p*s < 0.041), but not psychiatric traits (BF_{01}s > 3.40), were related to variability in opportunity costs for others. These findings show that when we decide to engage in prosocial behaviours depends on the environment and how prosocial and self-benefitting behaviours differ as environments change.

P2-C-80 - 'Mechanisms of prosociality in human social groups'

Paul Vanags [1], **Jo Cutler** [1], **Arkady Konovalov** [1], **Matthew Apps** [2], **Ian Apperly** [1], **Seongmin Park** [3], **Patricia Lockwood** [1]

[1] University of Birmingham, [2] Univesity of Oxford, [3] National Center for Scientific Research

Details

Objectives

Interacting in groups is a fundamental aspect of human social life. We regularly decide whether to incur personal costs to benefit a group. However, existing paradigms typically use two person dyads or lack ecological validity, and therefore do not reflect how groups interact in everyday life. Here we used social foraging theory to design a novel ecological task based on the 'producer-scrounger' model. In this model, producing generates rewards for the whole group, whereas scrounging relies on others producing to generate rewards. We tested whether producing behaviour would be higher when the relative cost was lower, and if people would produce more than the Nash equilibrium.

Methods

Participants interacted live online in groups of 6 (38 groups, n=228). On each trial they chose whether to produce resources for the group and incur a cost, or scrounge and free ride on others' production. A non-linear payoff matrix created a game-theoretic Nash equilibrium, that mirrored the diminishing returns of many natural resource settings. Importantly, as in real life, if no group member produced, everyone lost. We manipulated the relative cost of producing in 'summer' and 'winter' environments such that in summer producing was more costly than in winter. Finally, we measured individual differences in social and affective traits including to link them to behaviour within groups.

Results

As predicted, participants chose to produce more often in winter when the cost was lower ($t(405)=-10$, $p<.001$). Strikingly in both environments, they showed group-benefitting behaviour, defined as being higher than the Nash equilibrium ($t_{summer}(215)=7.9$, $p<.001$; $t_{winter}(215)=3.2$, $p=.002$). Intriguingly, the increase over Nash was larger in summer when costs were high than in winter ($t(215)=2.4$, $p=.020$). Models of choices over time showed that participants dynamically accounted for the choices of other players in their own decisions. Cluster analysis revealed 3 types of individual decision-making strategies: group benefitting, self-benefitting and mixed. Finally, variability in empathy ($r=-0.17$, $p=.014$), prosociality ($r=-0.25$, $p<.001$), and psychopathy ($r=0.16$, $p=.019$) were correlated with the difference in production of winter-summer. Those higher in interpersonal psychopathy were more willing to contribute to the group when the cost was low. In contrast, those high in empathy and prosociality were more group benefitting in high than low-cost environments.

Conclusions

These findings highlight surprising human group benefitting behaviour in ecological settings and reveal preliminary trait differences in social decision-making that can impact the behaviour of individuals in a group.

P2-D-81 - Sequential Information Preferences in Uncertain Decision-Making

Uma Karmarkar [1], **John Clithero** [2], **Carolina Raffaelli** [1]

[1] University of California, San Diego, [2] University of Oregon

Details

Objective: In uncertain choices, individuals can first attend to the nature of the possible outcome, or to the probability that outcome will occur. Prior research suggests that early information is more influential in shaping choice, and that being shown rewards before risks can increase a financial prospect's perceived value. Here we explore which information people select to learn about first *for themselves* and how this may influence choice.

Methods: The core task uses a lottery ticket with two sections: "$ amount you could win" and "% chance (probability) of winning." Participants learn that each ticket's information is randomly generated, and choose the section they want to reveal first. After this information is shown, they view the other section and asked willingness-to-pay (WTP) for the ticket.

Results: We find that individuals prefer to see risk before reward. In Study 1 a 58.4% majority selected probability first in the lottery task (n=402; p<.001). However, information choice did not influence WTP (p=.901). To unpack potential mechanisms, consider that although a lottery ticket implies a risky gamble, the task is ambiguous with unknown probabilities in addition to unknown outcomes. Thus Study 2 had participants (n=501) complete the lottery task and a separate measure of ambiguity aversion. Participants preferred viewing risk information first (65.7%, p<.001) and choice of risk was significantly associated with ambiguity aversion (logit; B=.043, p<.02). This was replicated in Study 3 with an expanded number of unknown information fields. Participants (n=580) saw a lottery ticket for a gift card where they could first view the amount of money, probability of winning, or retail store associated with the card. The proportion who chose risk first (46.7%) exceeded that for reward (37.9%, p<.001) or retailer (15.4%, p<.001; X^2(goodness of fit)= 91.17, p<.001). Ambiguity aversion again predicted choosing risk first (logit; B=.062; p<.001), but first information choice did not influence WTP (ps>.3). Study 4 thus tested whether risk-first choices might influence decision *processes* instead of value. After participants (n=375) chose to view risk or reward, they were asked if they wanted to continue (see the rest of the information), or exit (withdraw from the choice). The 59% majority (p<.001) who viewed the probability first exited the task at a higher frequency (X^2=42.5; p<.001).

Conclusions: People prefer viewing probability information before outcome information, which may be dependent on their aversion to ambiguity. While this self-selected sequence may not impact subjective value, the early attention to risk can cause decision withdrawal.

P2-C-82 - Developing a Computational Model for Unraveling Theory of Mind Dynamics in Social Interactions

Jan Gläscher [1], **University Medical Center Hamburg-Eppendorf (Uke)** [1]

[1] University Medical Center Hamburg-Eppendorf

Details

Theory of Mind (ToM) involves creating dynamic mental models to understand and predict others' behaviors. It extends beyond basic reasoning ("I think you will do X.") to recursive thinking ("I think that you think I will choose X."), also known as ToM Level-2. There is ongoing debate about which level of ToM is most suitable in various social contexts, with suggestions that competitive scenarios might require more complex ToM levels for strategizing against opponents. Similarly, cooperative interactions also demand advanced mental models for effective coordination. Additionally, it remains unclear how ToM adapts when individuals switch dynamically between different interaction types.

To explore this, we developed a novel Social Foraging Task (SFT) comprising four steps: patch decision, prediction, game decision, and outcome. In the patch decision, participants choose between patches with varying resource levels, which regrow if not selected. This choice determines the trial's context—

cooperative (Stag Hunt), competitive (Hide and Seek), or independent. Resource-dependent payoffs encourage switching patches, thus varying the interaction context. We used this setup to examine if Theory of Mind (ToM) adjustments occur when participants switch contexts. The SFT was tested on 74 pairs in an online experiment.

Model-free analysis of the patch decision revealed 3 principal strategies of players aiming for coordination: 1. **Resource Strategy**: both players select the patch with the largest resources. 2. **Adjusted Resource Strategy**: both players select the patch score with the largest adjusted resources (i.e. resource – switch costs). 3. **Preference Strategy**: some players selectively focused on one patch or neglected the other(s). We also observed a combination of these strategies. We are modeling the patch decision of the agent with a recursive ToM utility model that predicts the most likely action of the co-player by considering his observations of the task and his interactional preferences and derives the agent's action as a best response to this prediction.

In the game decision (choices in matrix games), we different distribution of strategies: during cooperation, participants predominantly selected the highest-reward option. Conversely, during competition, participant choice encompassed strategies like "win stay, lose shift", "tit-for-tat", and entropy maximization suggesting that competition elicits a heterogeneity of strategies. We are modeling the game decision with a new computational ToM model that estimates beliefs about the co-player strategies at Level-0 and computes best responses at higher levels. Using this different variant of this model, we are investigating ToM levels remained constant to change dynamically during context switches.

P2-D-83 - Stereoelectroencephalography (SEEG) Recording and Stimulation of Anterior Insula in an Instrumental Task with Frequent Outliers

Nina Sooter [1], **Peter Bossaerts** [2], **Fabienne Picard** [1]

[1] University of Geneva, [2] University of Cambridge

Details

In an exploratory study with SEEG, we investigated the role of Anterior Insula (AIns) in adaptation under uncertainty. We hypothesized two distinct potential roles. In line with traditional reinforcement learning, Hypothesis H1 states that AIns provides support for the reward system in the brain, tracking surprise to modulate the learning rate (e.g. Behrens et al., 2007). This modulation ensures that reward learning becomes approximately Bayesian (Nassar et al., 2010). Surprise coincides with salience, measured as the size of the prediction error and maximized upon outliers. We posit an alternative role, H2, whereby the AIns moderates the reward learning system, evaluating its outcomes against a generic reference model, and interfering when the performance is surprising compared to this model. Surprise is measured by autocovariance of prediction errors rather than their salience (Bossaerts, 2018). Engineers have used this approach, MRAC (Model-Reference Based Control), to enhance robustness in attempts to control unexpected changes in the environment (Nguyen, 2018). In (d'Acremont and Bossaerts, 2016)[AB], MRAC explained participants' choices significantly better than the Bayesian hypothesis.

We tested these hypotheses using an analogous task to [AB] combined with the SEEG recording of an epilepsy patient. The participant had to ensure that a randomly moving target remained centred while

subject to frequent outliers that either reverted (Treatment T, high uncertainty) or indicated a bias shift away from the centre (Treatment F, low uncertainty). The patient under-performed when outliers reverted (T) relative to when they did not (F), confirming [AB]. In a region of left AIns where previous fMRI analysis had revealed risk prediction errors (Preuschoff et al., 2008), overall SEEG activation in AIns under treatment T during the outlier trial was higher than under F despite equal overall salience, consistent with H2 and [AB]'s findings. Early activation (100-500ms) in both the outlier and post-outlier trial also aligned better with H2 than H1. Halfway during the outlier and post-outlier trials (500-1000ms), however, we recorded a second activation increase in both AIns and Amygdala that we intuit reflected salience (H1) rather than autocovariance (H2).

We then re-ran the experiment while stimulating the AIns for 2s during target moves. Half the stimulations were before outliers, the other half when the target move was small/absent. Choices moved closer to Bayes-optimal especially in T, where scope for improvement was highest. This is consistent with H2 provided stimulation attenuated the surprise signal.

Additional patients will be analyzed.

P2-D-85 - Behavioral and Neural Aspects of Knowledge-Based Effort Regulation – Introducing the Learning of Effort-Efficacy Task

David Stolz [1], **Nora Czekalla** [1], **Finn Luebber** [1], **Annalina Mayer** [1], **Laura Müller-Pinzler** [1], **Alexander Schröder** [1], **Yana Schwarze** [1], **Jan Weis** [1], **Sören Krach** [1], **Frieder Paulus** [1]

[1] University of Lübeck

Details

The key to navigating an uncertain world is to adjust one's effort based on knowledge about its likely consequences. Here, we present a novel task that allows to simultaneously assess these processes. In the Learning of Effort-Efficacy (LOEE) task, individuals learn about the irregularly changing conditions that determine whether their physical effort will be rewarded or punished. On each trial, they first choose between one of two buttons and may then press a hand-dynamometer to invest effort. For correct (or incorrect) choices they gain (or lose, respectively) points in proportion to their invested effort. On a subset of trials, they further rate how confident they are which button to choose next. In a first sample of subjects (n=48), we characterize their behavior in the LOEE-task. Subjects learn when to choose which button, and show systematic changes in choices, confidence, and effort investments depending on preceding outcomes. Further, they report changing confidence during the task and, akin to metacontrol, use their knowledge to guide the effort they invest. Further, we use model-based learning trajectories to show how computational (un-)certainty shapes task choices and compute a proxy of confidence that informs effort investments. Based on a second, independent sample (n=40), we replicate and extend these findings. Beyond showing the same basic behavioral patterns, we link computational certainty during choice, approximate confidence during effort investments, and processing of reward prediction errors to neural activity in the valuation system, salience system, and task-positive network. Together, the data from these two samples show that the LOEE-task allows to robustly and simultaneous assess mental and physical processes, as well as their neural correlates, that are key for goal-directed behavior. As such, the LOEE-

task provides a novel approach to studying these processes in an integrated fashion, which differs from previous experimental approaches that typically either focus on learning, or on effort regulation, but not their interplay. Although we here aim to characterize the task itself, it could potentially be applied in various clinical contexts. First, mental and physical aspects of fatigue are likely due to the action of immune mediators on dopaminergic neurotransmission. Second, depression is marked by beliefs that own actions are ineffective for bringing about desired outcomes. In summary, the LOEE-task could contribute to an enhanced understanding of neurocomputational involved in goal-directed behavior in general, but also in specific clinical domains.

P2-D-86 - An experimental and computational test of links between self-esteem, experience of (un-)certainty, and positive affect

Jan Weis [1], **David Stolz** [1], **Frieder Paulus** [1], **Sören Krach** [1]

[1] University of Lübeck

Details

Self-esteem (SE) is the evaluation of oneself as a person of worth, a central aspect of well-being and is closely tied to the experience of control and positive affect. While recent studies have shown how learning about others' approval of oneself shapes SE, less research has been done on performance-related processes using behavioral and computational methods. Using a novel experimental task, we investigate the roles of learning about control and of positive affective responses to task outcomes for state SE (SSE). Specifically, we hypothesized that both learning about the conditions for exerting control and positive affect relate to state SE. Participants (N = 74) performed a two-armed bandit learning task and earned money based on their performance. On each trial, participants invested effort using a hand-dynamometer and gained or lost points in proportion to that effort depending on whether the choice was correct or not. Positive affective responses to task outcomes were rated on a subset of trials and state SE was measured before and after the experiment. Further, subjects rated the certainty they experienced during the task in a post-experimental interview. As hypothesized, we find that positive affect and subjective certainty during the experiment were predictive of post-task performance-related SSE. Moreover, affective responses to trial outcomes were stronger for individuals with higher pre-task performance-related SSE. We further tested the association of behavioral data with computational parameters that formalize how subjects learned about the conditions for exerting control. Bayesian model comparison strongly supported a model architecture with volatility-informed learning rates (3- level Hierarchical Gaussian Filter, HGF; protected exceedance probability = 0.998) over models assuming constant learning (Rescorla-Wagner models, 2-level HGF). However, results of computational modelling remain inconclusive as some key parameters could not be recovered, whereas others were not associated with SSE. Together, our findings support the notion that beliefs of being in control over the environment and affective responses to performance-feedback are important factors for the development of SSE, which itself also shapes affective regulation. Further research combining behavioral and computational methods will contribute to our understanding of the development and formation of self-esteem, and its relevance to well-being.

P2-D-87 - The Pleasures of Suspense Enhance Risk-Taking

Maria Pitteri [1], **Roxane Morand** [1], **Bastien Blain** [1]

[1] Panthéon-Sorbonne University

Details

Objective

People invest considerable time in consuming non-instrumental information like stories in movies and books, which often feature two main elements: suspense and narrative distance. Suspense, or the uncertainty and its expected resolution, drives narratives (Baroni, 2016), while narrative distance—the perception that outcomes do not directly impact the observer—is inherent to them (Jose & Brewer, 1984). Although uncertainty is usually seen as aversive (Epstein, 1999), it can be enjoyable when it has no personal consequences. This study examines whether suspense is pleasurable and influences risk-taking behavior, and if these effects are more pronounced when participants are acting on behalf of others.

Methods

Experiment 1 utilized modified blackjack games to create suspenseful sequences (Li et al., 2021), based on a model that formalizes suspense as expected belief change (Ely et al., 2015). In each of 18 rounds, 5 cards were drawn from an initial deck. Each draw started with two preselected cards. Just before drawing one of them using a lottery, participants in the "suspense rating group" (n = 34) rated their level of suspense. If the cumulative suspense score exceeded 3 by the last round, they lost. The "satisfaction group" (n = 33) rated their satisfaction after the final draw. Participants played either for themselves (n = 32) or for others (n = 35) (“perspective”).

Experiment 2 allowed participants (n = 34) to choose their level of uncertainty, with a cumulative score below 3 needed to win. Each of the 5 rounds involved choosing from 10 lotteries which varied from round to round, but consistently followed the same structure in terms of expected value and orthogonalized variance. The starting point was manipulated to be neutral (0), favorable (-10), or unfavorable (10). Participants played either for themselves (n = 17) or for others (n = 17).

Results

Satisfaction ratings were more strongly associated with suspense in the third-person condition than the first-person condition ($t(30) = 2.12$, $p < .05$). When freely choosing suspense trajectories, participants won most of the time (>95%) but took risks in favorable situations, irrespective of the perspective ($t(32) = 3.0$, $p < .01$). However, participants did not seek to secure victory as quickly when playing for others compared to when playing for themselves ($t(32) = 2.37$, $p < 0.05$).

Conclusions

While uncertainty is thought to be aversive, this study suggests that uncertainty, in the form of suspense, is pleasurable, particularly when playing for others, and can lead to unnecessary risk-taking behavior.

P2-D-88 - Risk attitudes causally emerge from magnitude representations in intraparietal cortex: Evidence from combined TMS-fMRI

Gilles De Hollander [1], **Marius Moisa** [1], **Christian Ruff** [1]

[1] University of Zürich

Details

Risk attitudes describe the willingness to choose options that increase average returns but have more variable outcomes. We recently showed that individuals with a noisier number sense, defined as the precision of numerical receptive fields in the intraparietal cortex, are more risk-averse[1]. This link between magnitude-representation noise and risk appetite aligns with theoretical proposals that characterize individual risk attitudes as noise and biases in perception of payoff magnitudes[2,3]. Here, we establish the causal relevance of magnitude coding in parietal number fields for risk attitudes using fMRI-guided continuous theta-burst transcranial magnetic stimulation (cTBS) over the intraparietal cortex (IPS).

Seventy-eight participants took part in an fMRI experiment requiring them to decide between sure and risky options, using similar stimuli as in our previous study[1,4]. We fitted a numerical population receptive field model (nPRF)[1] to fMRI activity in the parietal cortex and selected 35 participants with the most robust numerosity fields for two additional behavioral sessions. In these sessions, rTMS was applied over either the individual nPRF cluster in IPS or the vertex (sham) in randomly-determined order.

IPS cTBS increased the proportion of risky choices, specifically when the safe option was presented first, rising from 56.3% to 61.6% ($F(1, 34)=5.00$, $p=0.032$). We fitted a computational model characterizing risk attitudes as noisy perception of numbers and prior attraction[4]. Model fits revealed that payoffs were represented with significantly more noise ($p=0.04$) after parietal rTMS. Moreover, a model using a more flexible mapping between magnitude and noisiness revealed that the representation of relatively low payoff magnitudes was affected most.

Neurally, IPS cTBS led to significantly lower nPRF amplitudes in the FMRI signal compared to vertex cTBS ($t(33) = 2.71$, $p=0.011$). Moreover, payoff decoding accuracy was significantly lower during IPS versus vertex cTBS (mean $r=0.142$ versus $r=0.092$, $F(1,34) = 4.99$, $p=0.032$), aligning with subjects' less consistent behavior.

Our study provides causal evidence that the precision of the parietal number sense determines individual risk attitudes. These findings substantiate neurocognitive theories of risk attitudes as biased and noisy perception and highlight the significance of perceptual processes and their neural substrates for economic choice theory.

1. Barretto-García, M. et al. Nat. Hum. Behav. (2023).
2. Khaw, M. W., et al. Rev Econ Stud (2020).
3. Vieider, F. M. Management Science (2024).
4. de Hollander et al. Presented at SNE 2023.

P2-D-89 - How the COVID-19 Pandemic Disrupted Optimism Bias in Belief Updating

Iraj Khalid [1], **Orphee Morlaas** [1], **Hugo Bottemanne** [1], **Lisa Thonon** [1], **Thomas Da Costa** [1], **Philippe Fossati** [1], **Liane Schmidt** [2]

[1] Paris Brain Institute, [2] National Institute of Health and Medical Research

Details

Optimistically biased belief updating is essential for mental health and resilience in adversity.

Objective: Here, we asked how experiencing the COVID-19 pandemic affected optimism biases in updating beliefs about the future.

Methods: One hundred and twenty-three participants estimated the risks of experiencing adverse future life events in the face of belief-disconfirming evidence either outside the pandemic (n=58) or during the pandemic (n=65).

Results: While belief updating was optimistically biased ($t(57) = 4.01$, $p < 0.001$, Cohen's $d = 0.53$) and Reinforcement-learning-like outside the pandemic (pxp = 1, Ef = 0.77), the bias faded ($t(64) = -1.06$, $p = 0.30$, Cohen's $d = -0.13$), and belief updating became more rational Bayesian-like during the pandemic (pxp = 0.90, Ef = 0.43; ß = -5.54, SE = 1.69, $t(232) = -3.28$, $p = 0.001$, 95% CI [-8.87 – -2.21]). This malleability of anticipating the future during the COVID-19 pandemic was further underpinned by a lower integration of positive belief-disconfirming information (ß = -0.03, SE = 0.02, $t(236) = -2.11$, $p = 0.04$, 95% CI [-0.07 – -0.002]), fewer ($t(121) = 2.40$, $p = 0.02$, Cohen's $d = 0.43$) but stronger negative outlooks ($t(121) = -3.01$, $p = 0.003$, Cohen's $d = -0.54$), and more confidence in base rates (ß = 14.11, SE = 4.52, $t(233) = 3.12$, $p = 0.002$, 95% CI [5.19 – 23.02]).

Conclusions: These findings offer a window into the putative cognitive mechanisms of belief updating during the COVID-19 pandemic, driven more by quantifying the uncertainty of the future than by the motivational salience of optimistic outlooks.

P2-D-90 - Subliminal risk influences valuation processes in ventromedial prefrontal cortex

Patricia Fernandes [1], **Sangil Lee** [2], **Joe Kable** [3], **Philipp Seidel** [4], **Jorge Almeida** [1], **Johan Eriksson** [5], **Bruno De Sousa** [1], **Fredrik Bergström** [1]

[1] University of Coimbra, [2] University of California, Berkeley, [3] University of Pennsylvania, [4] University of Regensburg, [5] Umeå University

Details

Little is known about the relationship between conscious awareness and value-based decision processes. Here we investigated whether subliminal risk could influence valuation processes by modulating conscious reward processes in terms of reaction time and neural pattern similarity in the anterior ventral striatum

(aVS) and ventromedial prefrontal cortex (vmPFC). Participants played a competitive and incentivized game with risky choice to accumulate points across trials in either a behavioral (n = 34) or an fMRI (n = 29) experiment. The game was a modified attentional-blink paradigm that rendered a "probability cue" (one of two symbols indicating 100% or 0% probability to win a risky reward) unseen. Following the probability cue, a choice between a safe (1 point with certainty) or a risky (>1 or 0 points depending on probability cue) choice had to be made. The risky amount varied between 2 and 5 points across trials. When probability cue was reported as unseen, value-maximizing choice (d') was not better than chance in either experiment. However, in the behavioral experiment there was a subliminal probability x conscious reward interaction ($F(1,33) = 10.23$, $p = 0.003$, $\eta^2 = .24$), such that reaction time was faster for larger than smaller risky rewards when probability was 100%, and the reverse when probability was 0%. In the fMRI experiment, there was no interaction but a main effect of subliminal probability ($F(1,28) = 6.59$, $p = 0.016$, $\eta^2 = .19$), such that reaction time was slower when probability was 100% than 0%. Moreover, it was possible to separately decode subliminal probability and conscious reward (all ps < .001) from neural patterns across the whole brain ($AUC_p = .73$, $AUC_r = .70$), in aVS ($AUC_p = .59$, $AUC_r = .60$), and in vmPFC ($AUC_p = .63$, $AUC_r = .59$). Importantly, it was also possible to cross-decode subliminal probability from conscious reward-related neural patterns in vmPFC (AUC = .56, $p < .001$), demonstrating similar neural patterns (while controlling for each other). The cross-decoding was mainly driven by vmPFC voxels with a positive relationship to reward amount as expected by subjective value signals. Taken together, these results suggest that subliminal risk can modulate valuation processes in vmPFC and influence behavior. Thus, conscious awareness is not necessary for information to be integrated into expected value in vmPFC. Moreover, the brain-wide contribution to probability decoding suggests information can have "global access" without conscious awareness, contrary to popular theories of consciousness.

P2-D-91 - Valuation reduces the description-experience gap by aligning probability distortions

Maike Brandt [1], **Stephan Nebe** [1], **Philippe Tobler** [1]

[1] University of Zürich

Details

Background: The description-experience gap (Barron & Erev, 2003; Hertwig et al., 2004; Weber et al., 2004) captures differences in economic decision making between situations where outcomes and probabilities are described versus learned from experience. However, the description-experience gap has primarily been studied with choices between two options, and it remains largely unclear whether the gap persists for preference elicitation methods assessing single options in isolation, such as Willingness-to-Accept (WTA). We therefore experimentally compared the description-experience gap for binary choice against ranked WTA.

Methods & Results: Using probability-matched decisions in the domain of gains, we replicated the description-experience gap for binary choice scenarios, both in the laboratory (N = 166) and online (N = 55) and with different sample numbers (10 and 20) and choice sets.By contrast, the gap did not manifest when the value of lotteries was determined with ranked WTA.

Applying Prospect Theory to our data, we found no differences in risk attitudes or probability distortions between description and experience for ranked WTA assessments ($MD_{riskAttitude} = -.051$ [-.140, .033];

$MD_{probabilityDistoriton}$ = -.023 [-.099, .048]). In both description and experience, individuals were moderately risk-averse and overweighted small probabilities. Conversely, in binary choice scenarios, individuals significantly underweighted small probabilities in experience compared to description (MD = -.459 [-.545, -.365]). In addition, individuals were significantly less risk averse in experience than description (MD = -.231 [-.306, -.152]).

Conclusion: Taken together, valuation of lotteries in isolation abolished the description-experience gap in risk attitude and probability distortion. These results indicate that the description-experience gap must arise primarily from the attribute- or option-level comparison processes underlying binary choice.

P2-D-92 - Meta-cognitive judgements of confidence are susceptible to momentary fluctuations in fatigue

Katarzyna Dudzikowska [1], **Nikita Mehta** [1], **Matthew Apps** [1]

[1] University of Birmingham

Details

Objective: As we engage in demanding tasks, we form metacognitive judgments of how confident we are to succeed. Classical theories suggest that confidence is a read out of the probability of successfully obtaining rewarding outcome. However, recent work has shown that effortful tasks can lead to sensation of fatigue that fluctuates on a moment-to-moment basis. This momentary fatigue is underpinned by two components, a recoverable component that increases when we exert higher effort, but decreases with rest, and an unrecoverable component that gradually increases during tasks but does not subside with rest. It has been suggested that confidence and fatigue are linked, as both require meta-cognitive judgements about internal states of the brain and body and are processed in similar systems in the brain. However, whether confidence fluctuates on a moment-to-moment basis in a similar manner to fatigue is unknown.

Methods: We developed a novel computational model in which confidence is susceptible to fluctuations in fatigue. We tested its predictions in two effort-based experiments where participants exerted prescribed levels of grip force (30, 39 and 48% of their maximum) on a dynamometer at percentages or took the same period of time (5s) resting. If they maintained the required level of force they obtained a random magnitude of reward between 10 and 80 points. Crucially, on each trial, they provided a subjective report of their prospective confidence in their ability to succeed at that effort level (experiment 1, n=41), or confidence and fatigue on interleaved trials (experiment 2, n=41).

Results: We find that despite participants maintaining high levels of success over trials (mean: 97.15%) their confidence in the ability to succeed significantly decreased over the course of the task ($F(1,65.9)=42.34$, $p<.001$) In line with our hypotheses, confidence changed on a trial-to-trial basis as a function of effort and rest ($F(3,6166)=14.41$, $p<.001$), mirroring fluctuations of fatigue. That is, even after succeeding, confidence decreased following mid-level and high effort (95%CI [-0.1, -0.03] and [-0.13, -0.07] respectively) and increased after rest (95%CI [0.01, 0.09]). Using formal model comparison, we show that our fatigue-based model of confidence can more accurately characterise trial-to-trial changes in confidence than alternative models based on previous performance.

Conclusion: Our results suggest that confidence judgements depend on moment-to-moment fluctuations in fatigue, and not just the probability of success. These findings point to shared mechanisms between fatigue and confidence that are known to influence economic choice.

P2-E-93 - Temporal discounting enlarges motivational difference of long-term goal and immediate gratification via striatal activities

Yang Xu [1], **Yachao Rong** [2], **Ping Wei** [1]

[1] Capital Normal University, [2] Henan Normal University

Details

Individuals with steeper temporal discounting (TD) tendency not only prefer immediate gratification than delayed gratification, but also achieve less for their long-term goals (i.e., pursuing delayed rewards). At the goal choice stage, TD may drive individuals to engage more in pursuing immediate rewards (IR), leading to procrastination for pursuing delayed rewards (DR). At the goal striving stage, however, large remains unknown whether and how TD tendency affects the motivation difference elicited by IR compared to DR. Here, we combined fMRI and a modified monetary incentive delay task (MIDT) to investigate the how the motivational value of IR and DR are represented and translated to action value in the brain, as well as the relationship with TD tendency. On each trial of the MIDT, we asked subjects (N=48) to discriminate the color of a target after a reward cue signaling the potential reward, IR, DR or no reward (NR). The response-contingent feedback then appeared to indicated whether the reward was obtained or not. Before scanning, subjects were informed that DR would be cashed out 3 months after experiment. After scanning, they were asked to finish a delay discounting task to measure TD tendency. In the MIDT, reaction times (RTs) are shorter when pursuing IR compared to DR. And RT difference in the MIDT, is positively correlated with the TD tendency (Fig.1). Neurally, both univariate and multivariate classification analysis indicate the mediating role of ventral striatum (VS) in the relationship between TD tendency and RT difference. More interestingly, we uncover that individuals with more similar TD tendency exhibit more similar neural pattern differences of VS when anticipating IR compared to DR. Using support vector regression, we further show that individual RT difference can also be decoded from these neural pattern differences. To solve the problem of reverse inference, we implemented a neural signature of reward, showing that reward pattern expression difference of IR versus DR indeed mediated the relationship between TD tendency and RT difference. Finally, we also uncovered a neural network for motor control, whose connectivity difference evoked by IR compared to DR support the behavioral difference (Fig.2). To summarize, our findings show that the motivational value differences of IR and DR are represented in the VS, as a function of TD tendency. And the increased communication of motor control network with VS may lead to larger performance difference. By focusing on the goal striving stage, we provide novel evidence accounting for the relationship between TD and goal achievement, and shed light on intervention of goal management.

P2-E-94 - Associations between duration memory and temporal discounting

Jonathan Grant [1], **Chloe Ott** [1], **Luis S. Garcia Campos** [1], **Karolina M. Lempert** [1]

[1] Adelphi University

Details

Objectives: Individuals vary substantially in their preferences for obtaining rewards sooner rather than later (i.e., temporal discounting), but the cognitive capacities underlying these individual differences remain elusive. It has been suggested that better memory abilities are associated with reduced temporal discounting, but which aspect of memory matters the most for intertemporal choice is unclear. We hypothesized that an individual's memory for how long experiences lasted (i.e., *duration* memory) is a critical factor underlying individual variation in temporal discounting. We examined if remembering time periods as shorter leads to an increased willingness to wait for delayed rewards. We assessed duration memory on three timescales: seconds, minutes, and days. In addition, as duration memory for an event depends on how many discrete events were perceived during the event, we also examined if individual differences in the processing of event boundaries underlie individual differences in duration memory and temporal discounting.

Methods: Participants (*n* = 64) completed one testing session, in which they first listened to an audio narrative while pupillometry data was collected. They then estimated the total duration of the audio narrative (this was our measure of minutes-level duration memory). Next, they did a duration memory task, in which they judged how far apart clips in the story seemed to be from each other (all clips were 1-min apart, so this reflects seconds-level duration memory). Then, they judged the lengths of previous time intervals in their own lives (e.g., "How long ago does May 3rd feel to you?") to obtain a measure of duration memory on the order of days (up to 90 days). Next, participants completed a standard intertemporal choice task before, finally, listening to the same audio narrative again, but this time pressing a key whenever they felt something new had occurred in the story (this was to determine subject-specific event boundaries). We estimated temporal discounting by fitting a hyperbolic model to choices from the intertemporal choice task, and we correlated temporal discounting with our three duration memory measures.

Results: Temporal discounting was not associated with duration memory on the time scale of seconds ($r = -0.010$; $p = 0.940$), minutes ($r = 0.198$; $p = 0.123$), or days ($r = -0.047$; $p = 0.714$).

Conclusions: Next steps include collecting data from middle-aged and older adults, as we expect that this will increase variation in duration memory, thus increasing statistical power. We also plan to relate pupil dilation evoked by event boundaries to both duration memory and temporal discounting.

P2-E-95 - A neuroeconomic explanation for contingency management

A. David Redish [1], **Anthony Defulio** [2], **Paul Regier** [3], **Rosemarie Davidson** [2], **Haily Traxler** [4], **Jenna Royle** [5], **Hannah Gass** [2]

[1] University of Minnesota, [2] Western Michigan University, [3] University of Pennsylvania, [4] University of Kentucky, [5] Eastern Michigan University

Details

Contingency management (CM) is one of the most effective treatments for substance use disorders. In CM, incentives (money, goods) are provided for objective verification of drug abstinence. Current theories of CM effectiveness are that it increases the cost of drug use by creating lost opportunity costs (the lost material incentive). While drug use is known to decrease with increased costs, drug use is notoriously inelastic. Further, the rewards provided in CM are small, and thus unlikely to produce the large effect sizes seen in typical CM studies. We will review the results of a published study (Regier and Redish 2015) and present a new, unpublished replication and extension (Davidson et al., under review). Study objective: To compare the observed effect sizes of CM against expected changes given known drug-cost demand curves. Methods used: Using a systematic review following PRISMA guidelines, we identified 79 studies that featured a mono-drug contingency CM intervention. Effect size was strongly correlated with incentive size (*r*=.42). We then examined cocaine-only (n=11) and nicotine-only (n=22) studies for further testing. We compared the observed effect sizes against the expected change in drug consumption obtained from the demand curves. Results: Consistent with the original findings of Regier and Redish, the cocaine studies showed a strong over-effectiveness, with 72% of the studies outperforming their predicted effects. Interestingly, the nicotine suppression studies did not differ significantly from the expected effectiveness (31% exceeded expectations). Conclusions: An alternate explanation for CM's success is that it shifts drug use from willingness-to-pay to choose-between questions, which may activate different decision-making processes. We will conclude with a discussion of the implications of these alternate explanations, including neurophysiological predictions, implications for identifying who will be best served by CM, as well as a discussion of why smoking cessation may be more pecuniary due to legality differences and the availability of over-the-counter options.

P2-E-96 - Taste processing is cognitively costly: an fMRI study on self-control under increased working memory load

Jakub Skałbania [1], **Marek Wypych** [2], **Andrzej Rynkiewicz** [3], **Łukasz Tanajewski** [1]

[1] Kozminski University, [2] Nencki Institute of Experimental Biology, [3] University of Warsaw

Details

Objectives

Dual-system theories propose that affective aspects of stimuli (e.g., taste in food-related decisions) can be processed rapidly and with negligible cognitive resources involved. This suggest that increased cognitive load should not affect taste processing. However, recent behavioral and neuroimaging studies have shown that a cognitively demanding task concurrent to food-related task may disturb taste processing, suggesting that the processing of affective aspects of stimuli requires cognitive resources. We explored this idea further through the lens of the tripartite neurocognitive model with the additional interoceptive system that regulates the activities of reflective and impulsive systems (involved in the cognitive and affective processing of stimuli, respectively). We asked whether and how the processing of taste in self-control task aimed at choosing a healthier over tastier food item requires cognitive resources. We hypothesized that the processing of both affective and cognitive components of self-control decision-making is disturbed under increased cognitive load. Specifically, we expected that the connections of the interoceptive system (represented by anterior insula (AI)) with reflective and

impulsive systems (represented by dorsolateral prefrontal cortex (DLPFC) and ventral striatum (VS), respectively) are reduced in high compared to low memory load conditions.

Methods

In an fMRI scanner, 49 participants made 120 food choices requiring self-control when keeping in memory 7- or 1-digit number in high (HL) and low (LL) working memory load conditions, respectively. Using the generalized Psychophysiological Interaction, we estimated the functional connectivity of AI with VS, as well as with DLPFC, and compared them between HL and LL conditions using one-sided paired T-tests.

Results

Despite no differences in the percentage of successful self-control between conditions, decision time was shorter in HL compared to LL condition. Connectivity between the right AI and right VS was greater in LL condition compared to HL condition (p=.037). Consistent with existing literature, we found a trend towards a stronger connectivity between right AI and DLPFC in LL compared to HL condition (p=.139 and p=.126 for right and left DLPFC, respectively).

Conclusions

Based on the tripartite neurocognitive model, we showed that the taste processing of food stimuli in a self-control task may require, at least to some degree, cognitive resources. In our ongoing studies we explore the practical implications of these findings, i.e., we ask whether, how, and under which conditions, a working memory task concurrent to food choice may result in improved self-control.

P2-E-97 - Delay discounting and willingness to wait in individuals with cocaine use disorder

Pazia Miller [1], **Anna Konova** [2], **Karolina Lempert** [1]

[1] Adelphi University, [2] Rutgers University

Details

Objective: People with cocaine use disorder (CUD) tend to show increased delay discounting, preferring smaller, sooner rewards over larger, delayed rewards more so than healthy controls. Although excessive delay discounting characterizes CUD, it is unclear if discount rate is a good predictor of relapse in individuals in recovery, since it does not capture persistence for rewards when there is temporal uncertainty. A new willingness-to-wait task measures how long people are willing to keep waiting for rewards that will arrive at an uncertain time. Here, we measured delay discounting and willingness to

wait in individuals with CUD in recovery, to see which task was more informative for predicting treatment outcomes. We expected that individuals with CUD would discount rewards more steeply than healthy controls (HCs), and that they would wait less time for delayed rewards compared to HCs, too. Methods: Sixteen individuals with CUD completed a delay discounting task and a willingness-to-wait task. On each trial of this task, participants waited through a random delay (0.2-40 s) for a 10¢ reward. They did not know how long the 10¢ would take to arrive, and they could quit waiting at any time and go to a new trial. Reward delays were distributed such that there was no optimal time at which to give up waiting. We plotted the probability that participants were still waiting for the reward at any given time within the 40 s delay period; the area under this curve (AUC) was the measure of willingness to wait, with higher values indicating longer wait times before quitting. We also assessed recent cocaine use and current cocaine craving. Results: To compare individuals with CUD to healthy controls, we took 100 random draws of 16 participants each from a dataset of 232 HCs who had completed the delay discounting and willingness-to-wait tasks. We conducted one-tailed t-tests to compare log-transformed discount rates between the CUD group and each of these 100 HC groups. The t-test was significant 70% of the time, showing that people with CUD had higher discount rates than HCs. We repeated this process for AUC and found no significant difference between groups (0% significant tests). In sum, we replicated previous research showing that individuals with CUD have higher discount rates than HCs do. Preliminary data suggests that willingness-to-wait is not lower in CUD compared to in HCs. Conclusion: Our results support the idea that delay discounting and willingness-to-wait are distinct aspects of impulsivity. Future analyses will attempt to link both constructs to relapse.

P2-E-98 - Increased delay discounting underlies behavioral symptoms in dementia and relates to medial pulvinar atrophy.

Anaïs Durand [1], **Richard Levy** [1], **Bénédicte Batrancourt** [1], **Liane Schmidt** [2], **Leonie Koban** [3], **Hilke Plassmann** [4], **Valérie Godefroy** [3]

[1] Paris Brain Institute, [2] National Institute of Health and Medical Research, [3] Centre national de la recherche scientifique (CNRS), [4] INSEAD Business School

Details

Behavioral variant frontotemporal dementia (bvFTD) is a neurodegenerative disorder characterized by core behavioral symptoms related to impulsivity. This condition also presents atrophy in brain regions important for valuation and computations such as trading-off smaller sooner (SS) vs. larger later (LL) rewards, coined delay discounting in behavioral economics. Our objectives were to: (1) investigate whether bvFTD patients show higher discounting rates than healthy controls (2) whether this altered discounting rate correlates with impulsivity-related bvFTD symptoms; (3) whether we can identify a structural brain network in which atrophy explains bvFTD's steeper discounting. BvFTD patients (N=24) and matched controls (N=18) performed two delay discounting tasks: one with money and one with food rewards. We compared discounting rates (log(k)) in bvFTD and controls and tested their correlations with bvFTD symptoms. We used participants' structural MRI data to perform whole-brain mediation analyses that investigate brain structures mediating the effect of bvFTD on the discounting of money and food. For both money and food rewards, the discounting rate was significantly higher in bvFTD patients than in healthy controls. Higher discounting of both money and food was associated with higher severity of the

symptoms of disinhibition and eating disorders. Whole-brain mediation analyses revealed that (1) lower grey matter density in left thalamic pulvinar, left parahippocampal cortex, and right temporal lobe predicted higher discounting of both money and food and (2) that grey matter density in these regions, including most prominently the medial pulvinar, mediated the effect of bvFTD on discounting. The impatience for immediate rewards measured by delay discounting might constitute a common mechanism of several core behavioral symptoms of bvFTD. Our study questions the central role of orbitofrontal cortex which has often been suggested to explain impulsivity in bvFTD and opens the way to further investigation of other additional brain regions such as the medial pulvinar. Future studies could investigate medial pulvinar structural damage as a transdiagnostic marker of impulsivity troubles and a potential target for their treatment.

P2-E-99 - Optimal and sub-optimal intertemporal decisions explain procrastination in a real-world task

Sahiti Chebolu [1], **Peter Dayan** [2]

[1] Max Planck Institute for Biological Cybernetics, [2] Max Planck Institute for Biological Cybernetics & University of Tübingen

Details

Objective: Procrastination is a universal phenomenon. It has long been recognized as a heterogeneous construct, with multiple psychological correlates and possibly many explanations. Despite this recognition, it is often reduced to the singular definition of defecting on an intended course of action, with temporal discounting as a primary explanation. Here, we asked if other mechanisms can explain procrastination in a real-world task (using rich data published by Zhang and Ma 2023).
Methods: Bachelor students in a psychology course (N=164) were required to participate in at least 7 hours of experiments in a 16-week semester, with each experiment lasting a minimum of 0.5 hours. Additional participation above 7 hours contributed to a small credit, up to 4 extra hours of work. We first used k-means (with a Euclidean metric) to cluster the patterns in which students distributed their efforts over time. We next modelled sequential decisions concerning the number of hours to work in each week, assuming this is determined by maximizing the expected sum of net future rewards. In our model, there are efforts and rewards associated with completing each ½-hour of work, efficacy governing the probability of successfully completion, and a small 'shirking' reward for doing an alternative task. We hypothesized the following mechanisms for procrastination and patterns of effort allocation: discounting of temporally distant rewards, a gap between real and assumed efficacy, non-linear increase of costs with amount of work, differential temporal discounting of rewards and efforts, and waiting for interesting experiments which might appear at some point in the future.
Results: k-means clustering of trajectories using the elbow method for identifying cluster number revealed 8 clusters, each capturing a different characteristic distribution of effort (Fig 1). Qualitatively, multiple models could replicate each pattern as evidenced by the distance between the simulated trajectories and each cluster. Quantitatively, we fit the models using maximum likelihood estimation (MLE) which again revealed that multiple models could fit each cluster (Fig 2; pseudo-R^2 = 0.15 to 0.35).

Conclusions: Our results suggest a taxonomy of different types of procrastination, from contradicting one's intentions to problematic patterns of intentional delays, along with preliminary evidence that multiple mechanisms along with temporal discounting could explain effort allocation in a real-world task, and offer alternative accounts for structurally different patterns of engagement. Our work is a contribution to understanding and characterizing the diversity of decisions underlying procrastination.

P2-F-100 - Dynamic Interplay of Hierarchical Rank and Social Contexts in Shaping Perceived Fairness

Yaner Su [1]

[1] University of Groningen

Details

Objective: Fairness is a fundamental principle of our social norms. Previous studies have shown that the social rank of individuals affects their perception of fairness. However, the effect of others' social rank on fairness norms is far from clear. Here, we aimed to examine the influence of social rank on perceived fairness in competitive versus cooperative contexts and the underlying neural mechanisms.

Methods: We conducted two experiments using the Ultimatum game (UG). Experiment 1 (N = 36) sought to provide behavioral evidence while Experiment 2 (N = 34) aimed to reveal event-related potentials (ERPs) responses. In both experiments, participants first learned about others' social rank through a competitive or cooperative game. Subsequently, they acted as a responder and decided to accept or reject different offers from proposers with different ranks in either competitive or cooperative contexts.

Results: Exp. 1 showed that participants accepted more offers from superior players than from inferior players ($p < 0.001$, $\eta_p^2 = 0.331$). Computational modeling revealed greater sensitivity to unfairness when offered by an inferior compared to a superior proposer in a cooperative context (95% HDI of $\alpha_{inferior}$ - $\alpha_{superior}$: [0.065, 1.587]). Exp. 2 replicated these behavioral findings and further revealed that when viewing a player's face, the cooperative context induced a greater P1 amplitude ($p = 0.029$, $\eta_p^2 = 0.15$), and high-ranking information induced a larger N170 response ($p = 0.04$, $\eta_p^2 = 0.14$). Moreover, sub-fair offers elicited a more pronounced FRN response ($p = 0.020$, $\eta_p^2 = 0.12$) while fair offers elicited larger LPP amplitudes than sub-fair offers ($p = 0.013$, Cohen's $d = 0.492$) in the superior rank condition.

Conclusions: This study provides behavioral and neural evidence that fairness perception is shaped by social ranks and differences between competitive and cooperative social contexts. It highlights the importance of considering social ranks combined with social contexts in studying fairness and provides valuable insights into how these factors influence human social interactions.

P2-F-101 - Neural correlates of reward anticipation after acute stress in alcohol use disorde

Yana Schwarze [1], **Johanna Voges** [2], **Alexander Schröder** [1], **Klaus Junghanns** [1], **Frieder Paulus** [1], **Sören Krach** [1], **Lena Rademacher** [1]

[1] University of Lübeck, [2] Universitätsmedizin Berlin

Details

Objective
Reward processing is altered after acute stress in the healthy population, characterized by increased activity in the striatum and amygdala and decreased activation in the medial prefrontal cortex (mPFC), which are key structures of the brain's reward circuit. In alcohol use disorder (AUD), evidence suggests dysregulation of the reward system as well as alterations of physiological stress systems, such as the hypothalamic-pituitary-adrenal axis and the autonomic nervous system. Despite the central role of stress and reward systems in AUD, there is a lack of research on the interplay between these. Thus, this study aims to investigate the effects of acute stress on reward processing in AUD.

Methods
Thirty-four individuals diagnosed with AUD and 34 healthy controls were recruited for a two-day study. Participants underwent acute stress induction using the Trier Social Stress Test and a control task on separate days. Cortisol samples, pulse rate, and questionnaire data on negative affect were collected before, (during), and after the tasks. Furthermore, participants completed a functional magnetic resonance imaging (fMRI) scan while performing an incentive delay paradigm with monetary and social rewards. A region of interest (ROI) analysis was conducted to analyze brain activity during the anticipation of the rewards, focusing on the amygdala, ventral striatum, and mPFC.

Results
For all stress parameters, interaction effects of group and stress were found: Pulse rate and cortisol responses to stress were blunted in AUD, whereas negative affect was increased. fMRI analyses revealed a significant group x stress interaction in the left amygdala for the anticipation of rewards: while controls showed increased activation under stress, it decreased in individuals with AUD. No significant effects were found in the other ROIs.

Conclusion
The data suggests alterations in physiological and affective stress responses in individuals with AUD as well as altered effects of acute stress on neural reward processing. This finding may reflect an interaction between stress and reward systems. Future studies are needed to understand the underlying mechanisms and the effects on motivation and decision-making. Understanding how alterations in reward processing following stress may influence decision-making could provide valuable insights into human behavior and choice.

P2-F-102 - Human Cerebellum and Social Navigation: Cerebello-cerebral Connections using Dynamic Causal Modelling

Meijia Li [1], **Frank Van Overwalle** [2]

[1] University of Birmingham, [2] Vrije Universiteit Brussel

Details

Background: Social navigation involves observing individuals navigate a socially rich environment sequentially while mentalizing their goals or preferences. The posterior cerebellum plays a central role in regulating human social behavior and guiding social navigation, yet the extent to which cerebello-cerebral pathways mediate this role remains insufficiently studied. The present study aims to investigate the effective connectivity between mentalizing areas of the cerebrum and cerebellum by applying a Dynamic Causal Modelling (DCM) analysis to four fMRI studies on sequential social navigation

Methods: We used fMRI data from four social navigation studies (*N* = 105) using similar task designs (i.e., navigation grid), protagonists (i.e., smurfs), and identical scanning procedures. Importantly, these studies showed activity in the mentalizing regions of the cerebellum and cerebrum. In this paradigm, participants were required to observe and memorize the movements in the trajectory of a social agent through a grid. As a control, the non-social condition was also implemented, using the same grid structure and trajectory but replacing the social agents with non-social objects such as pens, balls, and clouds.

Results: Results showed many closed loops (i.e., upward and downward connections with input and output in the same cortical area) of the cerebellar Crus 2 with cortical mentalizing areas, including the temporoparietal junction (TPJ), precuneus, and medial prefrontal cortex (mPFC), spanning both ipsilaterally and contralaterally. Connectivity patterns were only slightly modulated by experimental manipulations, involving (a) active encoding versus passive observation of navigational sequences, and (b) social versus non-social sequences. Furthermore, when participants observed behaviors defying social expectations or norms, upward connectivity from the right Crus 2 to the ventral mPFC, and downward connectivity from the precuneus to the left Crus 2 became stronger, while upward connectivity from the left Crus 2 to the ventral mPFC weakened.

Conclusions: Strong closed-loop connectivity (i.e., starting and ending in the same cortical regions) between brain structures is crucial in supporting the “sequencing hypothesis” of the social cerebellum (Leggio & Molinari, 2015), as it requires continuous information exchange between the cerebellum and the cerebrum. These results advance our understanding of the cerebellar role in effective connectivity and its changes in response to social violations during social navigation.

P2-F-103 - Neural structures underpinning ingroup bias and its malleability

Pyungwon Kang [1], **Yurong Sun** [2], **Juyoung Kim** [3], **Hackjin Kim** [3], **Sunhae Sul** [4], **Grit Hein** [5], **Philippe Tobler** [1]

[1] University of Zürich, [2] East China Normal University, [3] Korea University, [4] Pusan National University, [5] Wurzburg University

Details

Ingroup bias is the tendency to favor ingroup over outgroup members. It is pervasive in humans and contributes to inter-group conflict based on nationality or race. Most previous neuroimaging studies focused on functional differences between ingroup and outgroup conditions. Thus, it remained largely unknown whether structural differences also underpin ingroup bias. Here, we investigated whether grey-matter volume (GMV) is associated with implicit (implicit association test; IAT) and explicit (donation) measures of ingroup bias as well as their change.

In two studies, we measured the bias of South Koreans in South Korea against individuals from two different outgroups. More specifically, participants (Study 1: n=70, 35 female, age: 25.75±3.86; Study 2: n=108, 52 female, age: 23.44±2.24) decided whether to financially support charities benefitting the ingroup (Study 1 & 2: South-Koreans) or outgroup (Study 1: North Koreans in North Korea; Study 2: Southeast Asians in South Korea) and performed an IAT (Greenwald et al. 1998) associating the ingroup and outgroup with positive and negative words. We examined the association between GMV and behavior and whether they were specific or common for bias (implicit/explicit) and/or outgroup type.

We also investigated structural commonalities related to the change of implicit and explicit biases in different types of behaviors as well as different intergroup settings. Specifically, we measured IAT and donation before and after indirect exposure to outgroup-related information through a political event (Study 1: first summit of the two Koreas after 70 years, which occurred at the time of the first measurement) and to video-clips featuring interviews with Southeast Asians conducted in Korean (Study 2). Age, total intracranial volume, and study were controlled in all analyses.

We found that GMV in the putamen (Z=3.76, small volume corrected-SVC $p_{FWE} < 0.05$) positively correlated with explicit ingroup bias (donations). In contrast, GMV in ventromedial prefrontal (Z=4.74, whole brain corrected-WBC $p_{FWE} < 0.001$) and inferior temporal cortex (Z = 3.86, WBC $p_{FWE} < 0.001$) correlated positively with implicit ingroup bias, indicating that the GMV and ingroup bias relations are relatively task specific. In addition, GMV in a dorsal anterior cingulate cortex (dACC) cluster correlated with the change of both implicit and explicit biases in the two studies (IAT: Z= 3.93, WBC $p_{FWE} <0.05$, donation: Z = 3.09, SVC $p_{FWE} <0.05$) indicating that the dACC may generically reduce ingroup bias. Thus, both the strength of the ingroup bias and its malleability appear to be structurally underpinned.

P2-F-104 - How Does Soft Power Work? Behavioral Responses of Attraction-Based Influence

Aline Simonetti [1], **Hendrik Ohnesorge** [1], **Johannes Schultz** [1]

[1] University of Bonn

Details

Introduction: Soft power is defined as the ability to change the behavior of others based not on coercion or payments (hard power), but on attraction. It acts at the level of individuals and societies, being particularly important in international relations. However, contrary to hard power, the mechanism by which soft power takes effect is poorly understood. One possibility is that soft power acts as a kind of reward. This study tests the effect of a country's attractiveness on decision-making. Specifically, it investigates positive (i.e., attractive) and negative (i.e., repulsive) country soft power in several situations.

Methodology: This study comprised four tasks about 12 countries in a within-subjects design. First, participants rated how attracted/repulsed they were by each country (Country Assessment). Next, participants rated how much they liked 12 landscape photos (from 0="I hate it" to 100="I love it"), first without extra information, and then again with information about the photo's country of origin (Photo Task). Then, participants faced three hypothetical situations, each requiring them to rank the countries (from 1="most preferred option" to 12="least preferred option") in a certain context (Scenario Task). Finally, participants chose whether they would like a tip for solving a riddle from a helper from one of the countries they found attractive (riddle one) or from a country they found repulsive (riddle two) (Challenge Task). Participants could also decline a tip.

Results: We present preliminary descriptive results at the aggregate level (N=28). The Country Assessment allowed us to identify attractive and repulsive countries. In the Photo Task, liking ratings decreased after a photo's country of origin was revealed to be taken from a repulsive country (M_{pre}=82.4, M_{post}=70.2; $t(27)$=2.36, p=.026), while no change was observed with attractive countries (M_{pre}=79.4, M_{post}=79.7; $t(27)$=-0.19, p=.848). In the Scenario Task, repulsive countries were ranked lower across the three scenarios (M=9.2) than attractive countries (M=5.3) ($t(27)$=6.04, p<.001). In the Challenge Task, for those participants who wanted a tip, 88.9% chose a listed country when having attractive countries as options, while only 61.5% did so when having repulsive countries as options (p=.333, Fischer Exact Test).

Conclusions: These results suggest that a country's attractiveness influences decisions about the country, landscape photos and people, and thereby mimics the effect of soft power. The data also imply that negative soft power is more influential than positive soft power. These results are compatible with the idea that a country's soft power acts as a form of reward.

P2-F-105 - A meta-analysis of brain stimulation of the rDLPFC in social decision-making

Leticia Micheli [1], **Valerio Capraro** [2]

[1] Leiden University, [2] Università degli Studi di Milano-Bicocca

Details

Over the last years, several studies have attempted to causally examine the role of the right dorsolateral prefrontal cortex (rDLPFC) in social decision-making using brain stimulation techniques. The most prominent account for the role of the rDLPFC in prosocial behavior suggests that this brain region is involved in cognitive control and enforcement of prosocial behaviors. Nevertheless, empirical evidence is mixed, with some studies finding that inhibition of this region leads to more selfish behavior, while others find the opposite trend or non-significant results.

To systematize current findings regarding the involvement of the rDLPFC on social decisions, we conducted a random-effects meta-analysis of brain stimulation of the rDLPFC across different domains of social decision-making. Our inclusion criteria involved studies with healthy subjects and employing brain stimulation during a one-shot incentivized economic game to measure a range of social behaviors such as cooperation, altruistic giving, positive and negative reciprocity among others. An initial systematic review of the literature led to 174 articles. After careful examination of these articles based on our inclusion criteria, 38 reports were included in the meta-analyses and 84 effect-sizes were calculated and examined.

Despite substantial heterogeneity across studies, we find that stimulating the rDLPFC across different economic games makes individuals more likely to incur a cost to engage in social behaviors, compared to inhibiting the same brain area (overall effect size = -0.182, 95% CI = [-0.281, -0.083], z = -3.603, p < .001). No evidence of publication bias was found. By conducting a group meta-analysis and using the behavioral domain as a grouping variable, we find evidence of significant between-group heterogeneity (Cochran's Q = 16.98, df = 8, p = 0.030), suggesting that the effect of stimulating the rDLPFC indeed depends on the behavioral domain and it is primarily driven by studies measuring cooperative behavior (coeff = -0.757, z = -4.149, p < .001), positive reciprocity (coeff = -0.296, z = -2.685, p = 0.007), and negative reciprocity (coeff = -0.181, z = -2.148, p = 0.032). Interestingly, across all behavioral domains we observe the same direction of the effect, suggesting that the behavioral domain primarily influences the strength of the brain stimulation effect, rather than its direction.

Overall, findings suggest that when prosocial behavior is financially costly (as usually is the case in economic and incentivized games), the cognitive-control account for the role of the rDLPFC explains behavior across different domains of social decision-making.

P2-F-106, PS.02.03 - Psychological interventions increase motivation to exert effort to mitigate climate change

Jo Cutler [1], **Luis Contreras Huerta** [2], **Boryana Todorova** [3], **Jonas Nitschke** [3], **Katerina Michalaki** [4], **Lina Koppel** [5], **Theofilos Gkinopoulos** [6], **Todd Vogel** [1], **Claus Lamm** [3], **Daniel Västfjäll** [5], **Manos Tsakiris** [4], **Matthew Apps** [1], **Patricia Lockwood** [1]

[1] University of Birmingham, [2] University of Oxford, [3] University of Vienna, [4] Royal Holloway University of London, [5] Linköping University, [6] Jagiellonian University

Details

To limit the devastating effects of climate change, it is vital individuals engage in pro-environmental behaviours. Psychological interventions could be effective at promoting such actions. However, previous work has measured pro-environmental attitudes, missing a key aspect: effort. As pro-environmental actions are often more effortful, assessing the efficacy of interventions must include whether they increase willingness to exert effort for the environment. Here, we developed the novel Pro-Environmental Effort Task (PEET) and tested the effectiveness of 11 psychological interventions.

Methods

A large sample of participants (n=3,055), six representative samples from six countries across three continents, completed the PEET online. Participants chose to rest or 'work', rapidly clicking boxes on the screen, to earn rewards for a climate charity or a closely matched non-environmental cause. The effort required (50, 65, 80 or 95% of boxes clicked in calibration) and reward available (4, 12, or 20 credits) varied independently, enabling us to apply computational modelling to precisely quantify pro-environmental motivation.

Results

Strikingly, in the absence of an intervention participants were less motivated to exert effort to help the climate. Pro-environmental benefits were discounted by effort to a greater extent than donations to the non-environmental cause (OR=0.97, *p*=0.007). However, we found a small number of interventions increased motivation to benefit the climate, relative to the control charity. Specifically, interventions that targeted psychological distance and system justification significantly reduced the bias against helping the environment, across multiple measures and control analyses (*ps*<0.044). Computational modelling revealed these interventions affected devaluation of pro-environmental benefits by effort, independent of decision noise. Finally, while motivation to help both causes was associated with lower apathy and subjective effort, motivation to help the climate was specifically associated with belief in climate change and pro-environmental policy support (*ps*<0.001).

Conclusions

Our results highlight the importance of tasks with real effort and incentives to precisely measure motivation to engage in effortful pro-environmental actions. Using a novel task, we reveal the individual predictors of pro-environmental motivation and identify psychological interventions that promote relative willingness to exert effort to protect the planet.

P2-F-107 - An fMRI Study on Behavior and Neural Responses of Men to Female Body Odors

Maya Armin [1], **Elena Losse** [2], **Susanne Nehls** [2], **Thilo Kellermann** [2], **Natalia Chechko** [2], **Ute Habel** [3]

[1] Doctoral Candidate, [2] University Hospital RWTH Aachen, [3] Medical Faculty, Uniklinik RWTH Aachen

Details

Females of many species signal their fertile phase, as well as reproductive quality, to attract potential partners, and so, they influence the behavior of males (Mitchell et al., 2017; Coombes et al., 2018). In the pilot study of this project, body odors of ovulating (OV) women yielded increased activity in areas associated with facial processing and reward/motivation and the body odor of pregnant women (PRG) yielded activity in areas associated with empathy in men (Habel et al., 2021). In the current study, this was replicated while men were presented with axillary odors of menstruating (ME) women additionally. Samples were collected from 30 normally cycling (age: 19-34) and 30 pregnant women (age: 23-35, max. 16 weeks). In a single-blind fMRI experiment, 40 men (age: 20-34), were presented with the odors and air (NO) as control. All conditions were presented birhinally using a computer-controlled. Standardized portraits of female emotionally neutral faces of women were shown parallel to the odors.

All statistical thresholds were set at p<0.001, uncorrected, with a cluster-level correction (FWE p<0.05). Increased neural activity was found in response to ME compared to PRG in the left precentral gyrus and the right angular gyrus. Higher neural response was found during exposure to ME than OV in the bilateral medial temporal gyri, hippocampus, and cerebellum. OV>ME yielded no significant activity. No significant differences were found when OV was compared to PRG. NO>OV and ME>OV showed significant activity in the left medial temporal gyrus and a conjunction analysis of the mentioned contrasts revealed significant activity in this area.

The middle temporal gyrus has previously been shown to be involved in visual perception (Ishai et al., 2005). The similar activity pattern of this region in both, ME and the NO condition, could be because ME, similar to NO, has no reproductive relevance for the male participants. Thus, the reduced activity during the exposure to OV suggests that the images of faces without olfactory cues from sexually available women per se prompt activity in regions associated with emotion processing instead, as previously shown in processing faces (Kocsor et al., 2013).

P2-F-108 - Neurocomputational signatures of altered adaptive mentalization in autism

Niklas Buergi [1], **Arkady Konovalov** [2], **Caroline Biegel** [3], **Tania Villar De Araujo** [4], **Gökhan Aydogan** [5], **Christian Ruff** [5]

[1] Max Planck Institute for Biological Cybernetics, [2] University of Birmingham, [3] University Children's Hospital Zurich, [4] University of Zurich, [5] University of Zürich

Details

Objective: A key characteristic of autism is a limited ability to understand others' mental states. However, what specific neuro-cognitive mechanisms underlie these difficulties in mentalization remains poorly understood, and findings regarding neural correlates are highly mixed. Here, we identify a link between autistic traits and neurocomputational processes that dynamically adapt behavior to changing strategic thoughts of others.

Methods: Adult participants (47 autistic and 47 neurotypical matched for age, sex, and education) played repeated Rock-Paper-Scissors games against artificial opponents of varying mentalization depth (of the type "I think that you think that…"). The opponents were carefully calibrated in previous work to mimic human gameplay and were based on a new computational model that was also used to capture participants' strategizing (Buergi et al. 2024). This model assumes that agents try to infer the strategy of their opponent by mentally simulating what different opponents would play, and comparing these predictions to observed behavior. The resulting belief updates allow adapting to the reasoning process of the opponent and were robustly linked to a multivariate neural action pattern (Buergi et al. 2024). Here, we leverage this pattern as a normative benchmark for comparison between autistic and neurotypical participants.

Results: As preregistered, we found that autistic traits are linked to reduced performance in the game (standardized $\beta = -0.26$, $p = .005$). Model-based analyses suggest that this resulted from a decreased sensitivity to the way others reason ($\beta = -0.25$, $p = .020$), leading to reduced belief updates. Analysis of

concurrently acquired fMRI data did not reveal altered activation in any single area, but confirmed a reduced expression of the previously-identified neural pattern with higher autistic traits (β = 0.35, *p* = .003). Both reduced sensitivity and altered neural activation patterns mediated the effect of autistic traits on task performance, corroborating their role in explaining the observed differences.

Conclusions: We provide a mechanistic account of mentalizing difficulties in autistic people following a preregistered analysis protocol. Our results characterize a specific neurocomputational process that underlies mentalization ability, is implemented by distributed activity across the brain, and is changed in autism. Our findings inform theories about the neurocomputational basis of mentalization and may aid in the detection and assessment of corresponding behavioral problems.

P2-F-109 - The Neurocomputational Basis of Responsibility in Reparation-Giving

Elijah Galvan [1], **Jeroen van Baar** [2], **Luke Chang** [3], **Alan Sanfey** [1]

[1] Donders Institute, [2] Brown University, [3] Dartmouth College

Details

When faced with the decision to help a person in need, we often find that providing this help comes at a cost to ourself. Classic theories in social psychology place the notion of responsibility at the heart of this trade-off, suggesting that these feelings of obligation dictate when, and why, people are willing to help others who have been injured or wronged. However, it still remains unclear how the composition of group affects the responsibility we feel towards another person when we cause harm to them and, by extension, how likely we then are to engage in reparative action. Using a novel socioeconomic task which we termed the Reparations Game, we investigated how feelings of responsibility and reparation-giving changed as a function of group composition in a fMRI study. Here, we found evidence which clarifies the computational mechanisms underlying how collective responsibility diffuses over a group of people. Specifically, our adaptation of Latané's (1981) canonical Social Impact Theory – which we termed the Responsibility Stakes model – posits that responsibility grows based on the total number of witnesses and those who are accountable for causing the harm, but that the desire to engage in reparative action diffuses only among those who are accountable. This novel computational model outperformed alternative models in both self-report and behavioral data, providing converging evidence for our account of the diffusion of responsibility. We also found that the right Temporoparietal Junction and right Anterior Insula encoded the predictions for the Responsibility Stakes model, providing a neural basis for the computations of personal responsibility in group decision-making contexts.

P2-F-110 - Dorsomedial prefrontal cortex has a causal role in social influence

Zhilin Su [1], **Mona M. Garvert** [2], **Lei Zhang** [1], **Todd Vogel** [1], **Jo Cutler** [1], **Masud Husain** [3], **Sanjay Manohar** [3], **Patricia Lockwood** [1]

[1] University of Birmingham, [2] Julius-Maximilians-University Würzburg, [3] University of Oxford

Details

The medial prefrontal cortex (mPFC) has long been linked to processing social information and to economic decision-making. Several studies have suggested that dorsal portions of mPFC (dorsomedial prefrontal cortex, dmPFC) are involved in processing social information, while the ventral parts (ventromedial prefrontal cortex, vmPFC) may be relatively more specialised in processing information pertinent to the self. However, these conclusions have often been based on functional neuroimaging studies and therefore the necessity of these distinct subregions for social and economic processing remains unknown. Here, we tested a large sample of patients with rare focal damage to mPFC (*N*=33), those with lesions elsewhere (*N*=17), and healthy controls (*N*=71) (total *N*=121). Participants completed a temporal discounting task where they also learnt about the discounting preferences of two other people, one who was more impulsive and one more patient—before making their own discounting choices again. We used Bayesian computational models to estimate discounting and susceptibility to social influence and used these parameters in lesion symptom-mapping. We revealed that mPFC damage increased people's susceptibility to impulsive social influence compared to healthy controls, and increased overall susceptibility to social influence compared to those with lesions elsewhere. Importantly, voxel-based lesion-symptom mapping (VLSM), showed this heightened susceptibility to social influence was attributed specifically to damage to the dorsomedial prefrontal cortex (dmPFC, area 9). In addition, mPFC patients showed increased temporal impulsivity at baseline, with this heightened temporal impulsivity associated with lesions in the ventral portion of mPFC (area 13) as well as ventral striatum, in exploratory analyses. Therefore, we show that the dmPFC is causally implicated in susceptibility to social influence with distinct ventral portions of vmPFC involved in economic decision-making.

P2-F-111 - Characterizing BPD- and PTSD- feature-related variation in social trust learning

Damian Stanley [1], **Alexandra Voce** [2], **Karolina Lempert** [1], **Jack Grinband** [2], **Teresa Lopez-Castro** [3], **Erik Fertuck** [3]

[1] Adelphi University, [2] Columbia University, [3] City College of New York

Details

Knowing and evaluating who is trustworthy is an essential social cognitive capacity, and its impairment can have devastating consequences. Borderline Personality Disorder (BPD) and Post-Traumatic Stress Disorder (PTSD) have overlapping features and exhibit impairments in trustworthiness appraisal. Both are associated with severe functional impairments and high suicide rates. Previous work on trustworthiness appraisal in BPD and PTSD has focused on static, one-time judgments. However, evaluating others' trustworthiness is a dynamic process in which we incorporate new information and adjust our predictions. Here, we used a learning paradigm to investigate whether variation in social learning is associated with features of BPD and PTSD.

Across 2 blocks, 418 prolific participants (203m, 211f, 4o; 18-52yo) learned about 6 partners' trustworthiness in a multishot trust game (22 rounds/partner, 3 partners/block). Each round, participants saw their partner's picture (neutral expression) and decided how many points (1-5; self-

paced) to send to their partner, which then quadrupled in the transaction. Participants then saw whether their partner “shared” back half of the proceeds or “kept” them all. Partners had varying share response contingencies (share% = 80%, 65%, 35%, and 20%), which shifted once (by bias towards share/keep, level of certainty, or both; i.e. 35%->80%) between rounds 8 and 14. As incentive, earned points served as raffle tickets for a $50 bonus. In a matched nonsocial condition, participants learned about 6 lotteries. Multiple dimensions of psychological functioning relevant to BPD and PTSD were assessed.

Participant learning rates (LRs) and outcome biases (OBs; weighting of positive vs negative outcomes) were modeled with a modified Rescorla-Wagner update rule. LRs were significantly higher in the social (vs nonsocial) condition (95% CIs exclude 0). Negative outcomes carried more weight in both conditions, but more so in the social condition (95% CIs exclude 0). Individual differences in social vs nonsocial LRs correlated positively with participants’ scores on the Rejection Sensitivity Questionnaire ($r(294)=0.21$, $p<0.001$) and the Levels of Personality Functioning Scale ($r(294)=0.19$, $p<0.001$), which both assess personality functioning relevant to BPD. Notably, these correlations were driven by negative associations with nonsocial LRs rather than positive associations with social LRs. These findings highlight the complexity of variation in social learning ability and its relationship to psychological disorder, as well as the need for more nuanced research that delineates social from nonsocial learning and decision-making.

P2-F-112 - The Neural Signature of "Motive Cocktail" in Altruistic Behavior

Xiaoyan Wu [1], **Xiangjuan Ren** [2], **Chao Liu** [1], **Hang Zhang** [3]

[1] Beijing Normal University, [2] Max Planck Institute for Human Development, [3] Peking University

Details

Objective: Previous research has identified socioeconomic motives as essential for explaining altruistic behaviors, but most studies focus on only one or two motives, such as social equality and efficiency. This limited focus hinders their ability to explain the diversity in altruistic behaviors across contexts and individuals. In an earlier study (*N*=157 in the lab and *N*=1258 online), we demonstrated that third-party punishment and helping behaviors (i.e., penalizing norm transgressors or assisting victims at a personal cost) best aligned with a model featuring seven socioeconomic motives, termed a "motive cocktail". This study provides neural evidence for the co-existence of these motives in the same task.

Methods: We recorded Magnetoencephalography (MEG) signals from 38 participants while they performed a third-party intervention task. In each trial, participants, as unaffected third parties, observed a dictator ("transgressor") allocating more tokens to themselves than to a receiver ("victim"). The inequality information was displayed first, followed by the cost information on a separate screen, after which participants could choose to intervene by accepting a costly offer to punish the transgressor or help the victim, or to keep all their tokens. Across 400 trials, we varied the inequality (50:50 to 90:10), intervention cost (10 to 50 tokens), impact ratio (1.5 or 3.0), and scenario (punishment or helping), resulting in 100 conditions.

Results: As in our previous study, participants' choices to intervene were influenced by the four independent variables, showing both main effects and interactions; these choice patterns were best explained by the "motive cocktail" model. Regression analysis on the MEG time sequence revealed distinct spatial-temporal signatures (according to MEG sensor topography) for different variables and socioeconomic motives. Following the inequality window, inequality was first encoded in the parietal region. After the cost window, not only was the cost encoded, but also inequality was reactivated in the temporal region. The scenario signal persisted throughout the decision process. The model-based socioeconomic motives were mainly encoded in the frontal region, but with distinct temporal dynamics: self-interest emerged at an early stage, followed by two types of inequality-related representations and finally the inequality discounting motive (a compound motive suggesting a systematic discounting of others' inequalities with increasing costs).

Conclusions: The seven socioeconomic motives jointly shape third-party intervention decisions with distinct temporal dynamics in the human brain.

P2-F-113 - Emotion vs. Reward Prediction Errors in Social Decisions Across Adulthood

Colleen Frank [1], **Kendra Seaman** [1]

[1] University of Texas at Dallas

Details

OBJECTIVE: The difference between predicted and experienced rewards—i.e., reward prediction error (RPE)—has been shown to drive learning and guide future decisions. However, prediction about future rewards is not the only form of prospection used when people make choices. Predictions of future emotional experiences, sometimes known as affective forecasting, also drive decision making. A recent series of experiments found that emotion prediction errors (EPEs)—that is, violations of emotion, specifically valence, expectations—play an even larger role in punitive social decisions than RPEs (Heffner et al., 2021). However, this has not been tested in older adults, who make many highly consequential decisions and may rely more heavily on affect-based choice strategies compared to their younger counterparts. In this study, we tested the extent to which the relative impact of valence EPE and RPE on punitive choice decisions differ across adulthood. We predicted that valence EPE may be more important for choice in older age, whereas we predicted that RPE would not vary in its predictive value across adulthood.

METHODS: Participants across the adult lifespan (N = 272; ages 25-85) completed a social economic task (i.e., the Ultimatum game) with 20 age-matched computerized partners. For each trial, participants reported the offer they expected to receive from that partner (predicted reward), as well as how they predicted they would feel when given this offer (predicted emotions: valence and arousal). Following the offer (actual reward), participants also rated their experienced emotions (actual emotion). Prediction error was the calculated as the difference between expected and actual rewards and emotions.

RESULTS: To determine age-related differences in the role of valence EPE, arousal EPE, and RPE in punitive social decisions, we ran a mixed effect logistic regression. Across adulthood, valence EPEs (OR =

–2.40) were better than RPEs (OR = –1.24) at predicting punitive decisions across the lifespan, and there was no age-related interaction. In contrast to previous affective forecasting research, valence emotion predictions were less accurate in older age (B = –0.30), which is likely related to older participants predicting that partners would share more money with them than they actually did (B = –0.36).

CONCLUSIONS: These findings suggest that people across adulthood rely more heavily on valence EPEs than RPEs, at least for one type of punitive social decisions. Additionally, the novel negative association between age and valence prediction error suggests that age-related differences in the accuracy of predicting future feelings may be domain-dependent.

P2-G-114 - Age-dependent strategies in motivating effort towards deadlines from adolescence to adulthood

Janine Pfirrmann [1], **Todd Vogel** [1], **Andrea Pisauro** [2], **Tabitha Hogg** [1], **Margot Gueguen** [1], **Laura Rourke** [1], **Matthew Apps** [1], **Patricia Lockwood** [1]

[1] University of Birmingham, [2] University of Plymouth

Details

Objective: The ability to work effectively under pressure before a deadline is a defining aspect of educational success. Performance on educational outcomes have lifelong implications, predicting future earnings, mental well-being, and life satisfaction. However, meeting deadlines is effortful, and people typically find effort aversive. Yet, in some contexts, people put in great efforts for lower immediate reward, a phenomenon known as the effort paradox. Despite its developmental significance, we still lack a clear understanding of whether our motivation to exert effort before a deadline varies from adolescence to adulthood.

Methods: A large sample of participants aged 11-35 years (n=187, 52% females) completed a task where they had to motivate physical effort to reach a deadline. Participants engaged in a scientific mission on the moon to collect alien space rocks (rewards, converted into monetary payment), while also requiring physical effort (calibrated to each participant's maximum voluntary contraction) to charge their spaceship's battery to a specific threshold. On each trial, participants chose between a higher-effort option, which was associated with greater battery charging but lower reward, and a lower-effort option associated with higher reward. Importantly, failure to reach the charging threshold resulted in an inability to return to Earth and the loss of all collected rewards. We compared deadline pressure (work left to reach the threshold/time left), total rewards earned and trial the spaceship was launched across ages.

Results: Strikingly, younger adolescents were significantly more proactive, selecting higher effort in exchange for less reward on earlier trials (*OR=1.12 , p<.001*), thereby reaching the effort charging threshold earlier than adults (*r=0.33, p<.001*). Despite younger adolescents' higher motivation to reach the battery's threshold early, we found no age group differences in the number of spaceships launched (*X2 = 1.33, p=0.514*), nor the total amount of reward collected (*F=0.632, p=0.533*). Finally, we found that deadline pressure induced effort-seeking behaviour across age (*OR=4.12, p<.001*).

Conclusions: Our findings suggest that early adolescents use a less adaptive approach to allocating effort towards deadlines. They choose to exert greater effort at the beginning even when this strategy does not increase their rewards. Showing the effort paradox across all ages further indicates that effort can be rewarding in the presence of a goal. Together, our results have important implications for understanding effort processing and decision-making from adolescence to adulthood and ultimately shaping educational practice.

P2-G-115 - Characterizing the dynamics and experience of value-based decision-making in childhood and adolescence

Mahalia Prater Fahey [1], **Xiamin Leng** [1], **Megan Talikoff** [1], **Joonhwa Kim** [1], **Amitai Shenhav** [1]

[1] Brown University

Details

Aim: Despite recent efforts to systematically characterize the determinants, dynamics, and experience of adult decision-making, much less is known about decision-making earlier on. This limits our understanding of the development of relevant computations and circuits (e.g., related to evidence accumulation) and misses key periods during development characterized by increased agency and opportunity for engaging in value-based choice. Here, we seek to fill this gap by developing a novel paradigm for characterizing and comparing the process and experience of value-based choice from childhood through adulthood. We sought to test the extent to which recent findings in adults are present earlier in development, including 1) effects of relative and overall value on choice, 2) effects of option set properties on experienced conflict, and 3) effects of choice inclusivity (where selecting an option foregoes others) on both choice dynamics and experience.

Methods: Participants (ages 10-22 y.o.) played a developmentally appropriate decision-making game (current N=84). In the game, participants visited carnival tents where they were asked to choose from sets of items they had previously rated individually (e.g., games, books). Choice sets varied in relative and overall value and they were asked to choose their favorite option at each tent. At some tents they were then able to continue selecting as many items as they wanted (inclusive choices), whereas at others they were only able to choose that one (exclusive). After visiting all tents, they revisited each one and rated how much trouble they had making their choice.

Results: Consistent with previous adult findings, we found that both children and adolescents were faster ($ps<0.001$) and more likely to choose the best item first ($ps<0.001$) when relative value was

higher. They were also faster to choose an item when the overall value of the set was higher (ps<0.001). Both groups reported experiencing more choice conflict when the options were similar in value (ps<0.001) and when the overall value of the set was high or low (ps<0.001). Building on recent findings, we also found that both groups were faster (ps<0.04) and felt less conflicted when making inclusive choices (ps < 0.04). Most patterns also replicate in our young adults, suggesting our game is suitable for testing and contrasting with older ages.

Conclusions: Developmental work lags behind recent advances in understanding the dynamics and experience of value-based decision-making. Here, we connect these fields by developing a novel task which suggests even young children engage in similar processes and enables future work to further probe these processes.

P2-G-116, PS.02.01 - Age-dependent changes in hippocampal contributions to decision-making

Michael Cohen [1], **Karolina Lempert** [2], **Kameron Macnear** [3], **Frances Reckers** [4], **Laura Zaneski** [4], **David Wolk** [5], **Joe Kable** [4], **Camilla Van Geen** [4]

[1] University of Chicago, [2] Adelphi University, [3] University of Illinois - Urbana-Champaign, [4] University of Pennsylvania, [5] Hospital of the University of Pennsylvania

Details

This study investigates how changes in hippocampal-dependent processes during healthy aging affect different forms of decision-making. A sample of 86 participants (45 younger and 39 older adults) first underwent fMRI while they performed a choice task in which accurate performance relied on retrieving value information from a single past episode. The same group of people were then also scanned during a reinforcement learning (RL) task in which they learned the average value of four stimuli incrementally across repeated exposures. The traditional view of multiple memory systems posits that these two paradigms should rely on distinct neural substrates, with episodic memory in the first task recruiting the hippocampus and RL in the second task evoking neural activity in the striatum and ventromedial prefrontal cortex (O'Doherty et al., 2003; Squire & Zola, 1996). Given that we can simultaneously characterize age-related neural changes in either system, our experiment is well-positioned to test the validity of this distinction: we compare group-level performance on each of the two tasks, as well as the neural activity that underlies it, in order to measure whether age-related cognitive impairments preferentially affect one function or the other. We find that older adults were especially impaired at making decisions from episodic memory (age*reward interaction: b = 0.41, p = 0.016), and showed reduced activity in hippocampus and medial temporal lobe at encoding. In the reinforcement learning task, older adults accurately learned the value of the four stimuli (age*learning block interaction: b = 0.004, p = 0.84), and showed no differences in reward prediction error signaling in the striatum. Nonetheless, older adults did show subtle performance deficits in RL, especially at the end of learning (age difference in proportion correct choice on the final block: t = -1.86; p = 0.06). Using representational similarity analysis to assess how the hippocampus tracks distinct stimulus identities, we can link these subtle behavioral impairments to disrupted patterns of stimulus-specific neural activity. Together, these findings show that changes in

hippocampal processing with healthy aging underlie both larger impairments in decisions based on episodic memory and more subtle impairments in incremental learning.

P2-H-117 - Integrating habit and reinforcement learning within a classical drift-diffusion model

Viktor Timokhov [1], **Hugo Fluhr** [1], **Philippe Tobler** [2], **Stephan Nebe** [1, 2]

[1] University of Zurich, [2] University of Zürich

Details

Traditionally, habits are seen as stimulus-response (S-R) associations, insensitive to values. Miller, Shenhav, and Ludvig (2019) modeled choices by combining a habitual controller, sensitive only to action history, with reinforcement learning (RL), capturing goal-directed behavior. Zhang et al. (2024) extended this with sequential sampling by assuming that (1) drift-rate (i.e., slope of evidence accumulation process) depends on learned value difference (as tested in Pedersen et al., 2017), and (2) response bias (i.e., shift of the starting point for evidence accumulation) depends on the difference in habit strengths between options. They used RL to compute option values, or Q-values, and the habit learning model (Miller et al., 2019) to compute habit strengths, or H-values. But because of its complexity and lack of an explicit likelihood function, their model is hard to apply to empirical data, while the second assumption was not explicitly tested.

First, we validated the difference-based response bias assumption with a classical drift-diffusion model (DDM) using data of 220 participants performing the Rewards Pairs task (Nebe et al., 2024) - an instrumental learning task that manipulates the values of options and the frequencies with which they are chosen during learning. We focused only on test trials with equivalent values but different choice history, to compute habit strengths in choice, reaction time, and stimulus liking. DDM parameters were fitted with the *hddm* 0.9.8 package (individual/group-level estimations for response bias, other parameters at the group-level only, 6 chains, 50000 samples, 3500 burn-in samples). Response biases in the classical DDM correlated with choice-based habit strength (Pearson bootstrapped correlation, r=0.58, 95% CI [0.45, 0.66]), to a lesser extent with reaction time-based habit strength (r=0.39, 95% CI [0.23, 0.51]), and not with liking-based habit strength. These findings are compatible with the empirical data and support the assumption on response bias.

Next, we developed an alternative modification of the standard DDM. It retains the difference-based assumptions for drift-rate and response bias but uses fewer parameters (addressing the complexity challenge). Specifically, drift-rate is the scaled difference between Q-values (learned via the Rescorla-Wagner model) and response bias is the difference between H-values (learned via habit learning model). Moreover, it incorporates a likelihood function calculated via Wiener first-passage time distribution. Simulations demonstrated that this model qualitatively reproduces behavioral results with habit measures and accounts for individual differences in habit strength.

P2-H-118 - Neural Correlates of Choice Frequency in Habitual Behavior: An fMRI Study

Hugo Fluhr [1], **Viktor Timokhov** [1], **Philippe Tobler** [2], **Stephan Nebe** [2]

[1] University of Zurich, [2] University of Zürich

Details

Background: Habitual behaviors are pervasive in daily life. Prior research from our group (Nebe et al. 2024) indicates that past choice frequency can influence habitual behavior, yet neural underpinnings of this mechanism remain unexplored. Our ongoing, pre-registered, fMRI study aims to fill this gap.

Objective: This study aims to replicate previously obtained behavioral results of the role of previous choice frequency during an instrumental learning task. We further extend our previous work by identifying the neural correlates of previous choice frequency.

Methods: We adapted the "Reward Pairs" task from (Nebe et al 2024) for fMRI (longer intertrial intervals etc.). 71 participants were scanned while performing this task that involves learning stimulus-response associations through trial-and-error under time pressure, without salient outcome manipulations. Computational models combining reinforcement learning (RL) and a choice kernel (CK) are used to explain choice behavior.

Hypotheses: H1 Modeling Behavior: Participants' choice behavior is best explained by a model that integrates RL and CK, outperforming models with either component alone. **H2 Influence of Choice Frequency:** Previous choice frequency significantly influences choice behavior in a generalized linear mixed effects model. **H3 Neural Correlates:** BOLD responses in a priori regions of interest (e.g., ventral striatum, vmPFC for RL; dorsolateral striatum, parietal and (pre-)motor cortex for CK) correlate with RL and CK values during task performance. Integrated value signals (RL and CK) may also be observed in the ventral striatum and parietal cortex.

Results and Analysis Plan: H1: All models were fitted on the training data, the RL+CK outperformed all other models when predicting the choices in the test phase (BIC_{RL+CK}=311.74 vs BIC_{RL}=497.48 and BIC_{CK}=782.01). **H2:** Previous choice frequency significantly influenced choice behavior in a GLME (z=2.875, p=0.004). **H3** For neural data, trial-by-trial RL and CK values will be correlated with BOLD responses using regions of interest from previous meta-analyses and whole-brain exploratory analyses. This analysis aims to validate our computational models and uncover the neural mechanisms underlying the observed behavioral patterns. Currently we are conducting the analysis for H3.

Conclusion: By building on previous work and adding an imaging component to it, this study replicated previous findings on choice frequency and will elucidate the neural processes that support this behavior. Understanding these mechanisms could inform strategies for behavioral interventions in various contexts, including psychopathology and behavior change initiatives.

P2-H-119 - Distinct effects of anxiety and depression on reinforcement learning

Xinru Huang [1]

[1] Peking University

Details

Objective: The comorbidity and heterogeneity of depression and anxiety has long been a topic under discussion. The current study aims to reconcile results from previous studies by proposing that depression and anxiety, despite their comorbidities, are independently associated with learning in a standard two-arm bandit task.

Methods: We administered a reinforcement learning task and psychiatric questionnaires to a large subclinical subject sample through online experiment (190 participants). The reinforcement learning task included gain and loss conditions, each of which was composed of 100 repeated trials of binary choice between two slot machines with slowly changing reward probabilities. We employed RW learning models with positive and negative learning rates (bounded between [0, 1]), initial values (bounded between [-1, 0] in the loss condition, [0, 1] in the gain condition), perseverance (bounded between [-5, 5]), and inverse temperature (bounded between [0, 20]). For psychiatric questionnaires, the Self-rating Depression Scale (SDS) and the State-Trait Anxiety Inventory-Trait version (STAI) were used to measure symptoms of depression and anxiety respectively.

Results: We regressed individual learning rates on the scores of SDS and STAI. Depression and anxiety scores exhibited dissociable effects on learning rate: the depression score was negatively associated with the learning rate, whereas the anxiety score was positively associated with learning rates. This pattern remained stable for both positive and negative learning rates, and in both gain and loss conditions. Considering the high comorbidity of depression and anxiety, we implemented factor analysis on the 40 items of the two scales to reduce dimensions. Exploratory factor analysis revealed three factors, corresponding to a general factor loading on both SDS and STAI items, a depression-specific factor loading mainly on SDS items, and an anxiety-specific factor loading mainly on STAI items. Multiple regression analyses confirmed that the depression-specific factor was negatively associated with learning rate, while the anxiety-specific factor was positively associated with learning rate. A further check on the items indicated that items loading on the depression-specific factor mostly related to negative valence, while items loading on the anxiety-specific factor mostly related to low arousal.

Conclusion: We showed that the depression and anxiety traits tended to correlate with learning rates with different directions.

P2-H-120 - Mixed evidence for rational counterfactual learning

Kate Nussenbaum [1], **Nathaniel Daw** [1]

[1] Princeton University

Details

Objective: People learn from the outcomes of their own actions and from observing the counterfactual (cf) outcomes of actions they *could have* taken. In most studies of cf learning, participants observe a single, alternative outcome. Here, we asked whether people rationally prioritize learning to improve their future choices when faced with multiple cf alternatives. We hypothesized that participants would prioritize

learning from negative outcomes of favored options and positive outcomes of disfavored options, because such outcomes would indicate the need to update decision policies.

Methods: 30 adults recruited from Prolific completed 200 trials of a three-armed bandit task. On 140 free-choice trials, participants chose between three options that yielded binary rewards with independent, changing probabilities. On 60 randomly interspersed forced-choice trials, they were forced to select a specific bandit. The outcomes of all three options were presented on every trial.

We fit participants' choices with a reinforcement-learning model with eight cf learning rates that characterized how they learned on free versus forced-choice trials, from positive versus negative outcomes, and depending on whether the outcome of their selection was positive or negative. On free-choice trials, we expected that participants would choose their favored bandit, and learn more from positive vs. negative cf outcomes from the unfavored bandits, because positive outcomes indicate that it may be better to switch to the alternative option. On forced-choice trials, when participants could not always choose their favored bandit, we expected that upward cf updating would be attenuated — negative cf outcomes associated with their favored bandit would also indicate the need to update their decision policy, and thus may be prioritized.

Results: On free-choice trials, participants learned more from cf outcomes when their favored option did not yield reward, $F(1, 87) = 63.8$, $p < .001$. While we expected that participants would show upward cf updating, they learned more from negative vs. positive cf outcomes, $F(1, 87) = 134.0$, $p < .001$, in line with a confirmation bias. Further, these effects were attenuated on forced-choice trials, meaning we did not see greater downward cf updating as we had hypothesized ($ps > .056$).

Conclusions: We observed mixed evidence for 'rational' cf updating. Participants upregulated cf learning when doing so was likely to be most useful. However, they did not prioritize learning about the best alternative. Future experiments will explore why some choice contexts yield confirmatory, downward cf updating, while others yield more adaptive, upward cf learning.

P2-H-121 - Confirmation bias exists in the face of false information

Hamid Razi [1], **Thomas Sambrook** [2], **Neil Garrett** [2]

[1] PhD Student, [2] University of East Anglia

Details

Objective: Confirmation bias maintains unrealistic and polarising beliefs and impacts decisions across a range of domains from finance to politics to science. The mechanism that generates this bias is a learning asymmetry in which learning from confirmatory (α_C) exceeds disconfirmatory (α_D) information. A pressing question is whether this mechanism persists in the face of false information where the potential impact on individuals and society is far greater. We test this using a new economic decision-making task in conjunction with computational modelling in two separate studies.

Methods: Participants (study 1: N=47; study 2: N=57) made choices between pairs of options (abstract symbols). Following feedback (win/lose money), cues indicated whether the feedback was genuine or

false. Trial-by-trial choices were fitted to computational models which varied the number of learning rates. This enabled us to test (using maximum a posteriori probability estimation) whether updating was greater from confirmatory relative to disconfirmatory prediction errors (PEs) for inaccurate and/or accurate information.

Results: Model free analysis of choices revealed participants updated beliefs following inaccurate information (study 1: $t(46)=2.64$, $p=0.01$; study 2: $t(56)=3.47$, $p<0.01$) but at a reduced rate relative to accurate information (study 1: $t(56)=5.42$, $p<0.01$; study 2: $t(56)=5.48$, $p<0.01$) suggesting sensitivity to the accuracy cues. Next, we inferred the underlying learning processes using computational modelling. The winning model (exceedance probability > 99% in both studies) had 4 learning rates ($\alpha_{C,TRUE}$, $\alpha_{D,TRUE}$, $\alpha_{C,FALSE}$, $\alpha_{D,FALSE}$) plus softmax, perseverance and choice trace parameters. This model accounted for participant choices to a greater degree than models that (via fewer learning rates) failed to enable separate learning from confirmatory/disconfirmatory PEs or from accurate/inaccurate PEs. Examining the pattern of learning rates from the winning model revealed a robust confirmation bias following false ($\alpha_{C,FALSE}>\alpha_{D,FALSE}$) information (study 1: $t(46)=3.49$, $p=0.001$; study 2: $t(56)=5.74$, $p<0.01$) as well as following accurate ($\alpha_{C,TRUE}>\alpha_{D,TRUE}$) information (study 1: $t(46)=2.58$, $p=0.01$; study 2: $t(56)=3.91$, $p<0.01$).

Conclusions: Together, these findings suggest for the first time that the same mechanism known to give rise to confirmation bias from accurate information – a learning asymmetry – also exists in response to false information in simple instrumental learning settings. This serves to highlight how deep seated the bias is and poses timely questions as to how we should act to curb it in cases where it has the potential for harmful societal consequences.

P2-I-122 - Cardiac activity predicts monetary donations to charity videos

Anna Shepelenko [1], **Anna Shepelenko** [2], **Marina Tiuleneva** [3], **Anastasia Obukhova** [4], **Vladimir Kosonogov** [1], **Victoria Moiseeva** [1], **Iiro Jaaskelainen** [5], **Hanna Keren** [6], **Anna Shestakova** [1]

[1] National Research University Higher School of Economics, [2] Institute for Cognitive Neuroscience, HSE University, Moscow, Russia, [3] University of Trento, [4] Moscow Institute of Physics and Technology, [5] Aalto University, [6] Bar-Ilan University

Details

Conscious narrative processing can affect cardiac activity. Specifically, identical storytelling content evoke similar perceptual and cognitive processes among individuals, leading to synchronization of heartbeats. Recent evidence shows that reduced attention to narratives can result in decreased inter-subject correlation of heart rate (ISC-HR), which indicates the level of engagement of the individuals. Furthermore, attention focus on storytelling stimuli can cause heart rate deceleration (HR deceleration). Therefore, ISC-HR and HR deceleration could be used as indicators of the appeal of audio and visual materials, including in the context of evaluating advertising effectiveness.

The aim of this study was to investigate the relationship between heart rate while viewing a charity appeal and willingness to donate. Fifty-four healthy participants (N = 54) took part in the laboratory experiment. During the experimental procedure participants were presented eight three-minute videos, four of which

were fundraising ones. Following the viewing of each charity video, participants were given the opportunity to donate from a pre-received reward.

Our findings indicate that ISC-HR is positively correlated with donation size *rrm*(158) = .25, *p* = .002. In other words, that participants donate more when their ISC-HR increased while watching a charity video. Additionally, we found that willingness to donate correlated negatively with the mean of heart rate changes, *rrm*(80) = -.26, *p* = .016. This suggests that the greater the decrease in heart rate, the more likely participants were to donate a larger amount.

Thus, HR deceleration and heart rate synchronization during the conscious processing of charity appeals are associated with willingness to donate. These results may be useful for non-profit organizations to evaluate the effectiveness of charity advertising. In the future, we aim to study brain activity synchronization during conscious processing of these appeals using functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) methods.

P2-I-123 - Self induced framing to aid decision-making

Marc-Lluís Vives [1], **Pablo León-Villagrá** [2]

[1] Leiden University, [2] Brown University

Details

Decision frames influence how people act. These frames and the resulting decisions can be changed by manipulating how a problem is described. Here, we ask if people themselves can induce frame changes when thinking about a problem and how these frame changes affect decision-making and choice satisfaction. In our experiment, participants (N > 700)generated as many options as they would like for day to day scenarios as choosing a costume for a party or finding a gift for a friend. Then, participants selected one of the options they generated and reported their choice satisfaction. We found that choice satisfaction was higher when the option selected was more semantically dissimilar to the rest of the option set. We argue that this suggests that participants use a novel strategy to facilitate decision-making: Participants aimed to construct decision frames by generating options sets with a uniquely dissimilar option, which facilitated choice and increased satisfaction.

P2-I-124 - Relative Value Biases in Large Language Models

William Hayes [1], **Nicolas Yax** [2], **Stefano Palminteri** [3]

[1] Binghamton University, [2] Ecole Normale Superieure, [3] École Normale Supérieure

Details

Objectives

Studies of reinforcement learning in humans and animals have demonstrated a preference for options that yielded relatively better outcomes in the past, even when those options are associated with lower absolute reward. In contrast, while Large Language Models (LLMs) exhibit impressive reasoning abilities akin to human intelligence, the processes guiding their behaviour remain unclear. This study examines how LLMs approach reinforcement learning tasks to see whether these models would exhibit similar biases.

Methods

We had two LLMs, GPT-4 and Llama-2-70B make repeated choices between pairs of options with the goal of maximising payoffs. The task structure is adapted from (Hayes & Wedell, 2023) involving 8 options presented in 4 fixed pairs of increasing reward magnitude during the learning phase. Each pair is shown 5 times during training. Then, in the transfer phase, the agent chooses the best option from all possible combinations of options. While a rational agent would select the most rewarding option in every combination, humans tend to prefer an option that was locally optimal in the learning phase to another that wasn't optimal even if the preferred option gives lower absolute reward. Each LLM completed the task 30 times with default temperature, following a methodology similar to previous studies (Binz & Schulz, 2023), with a context explanation, outcome record, and choice prompt.

Results

Models engage consistently with the task and are near optimal at the end of the training phase. We evaluated the rationality of the models' behaviour in the transfer phase with a linear contrast on the choice rates, where negative values of the contrast indicate relative value encoding. GPT-4's choices led to a value of -0.17 ($p<0.001$, $d=1.37$) showing significant relative value encoding behaviour while Llama's behaviour did not show significant relative value encoding with a value of -0.02 ($d=0.12$).

Conclusions

While LLMs are known to replicate behaviour seen during training, bandit tasks are not common in a textual format meaning they are very unlikely to have been trained to replicate this relative behaviour. Finding biases akin to humans on such tasks in one language model and not in another one is surprising and raises questions about how these biases appear and what is their role in these models. This opens the way for further research into the origins and implications of economic biases in LLMs, humans and intelligence in general.

P2-I-125 - The common root of creativity-related preferences across domains.

Gino Battistello [1], **Sarah Moreno-Rodriguez** [1], **Emmanuelle Volle** [1], **Alizée Lopez-Persem** [1]

[1] Paris Brain Institute

Details

Decision-making is a daily and necessary practice for human beings, and creative decision-making is one of them. Previous research on creativity has proposed a breakdown of creative ideation into two main

processes: a generative phase for spontaneous idea associations and an evaluative phase for monitoring these ideas. During the evaluation stage, ideas are judged based on their adequacy and originality, two defining dimensions of creativity. Recent findings decompose the evaluation phase into monitoring and valuation. The subjective value of ideas (valuation) was found to be computed based on their adequacy and originality (monitoring).

The aim of the current study is to ascertain the consistency of preferences across diverse creativity domains. To achieve this, seventy-three participants engaged in free-generation tasks across three creativity domains: semantic associations, alternative uses, and drawings. Subsequently, they rated their responses based on likability (subjective value measurement) and perceived adequacy and originality.

Participants tended to provide ideas more rapidly when they liked them more across all three domains. This result emphasizes the motivational role of valuation in the creative process. Furthermore, across these domains, the likability of ideas stemmed from a combination of their adequacy and originality. Through computational modelling, we established that a similar non-linear value function, with consistent weighting and convexity parameters across domains, governed ideas' judgments. These findings reflect both the inherent nature of the valuation process observed in value-based decisions and contribute to consolidating our understanding of creative ideation.

P2-I-126 - Urgency and strategy selection in multi-attribute choice

Krzysztof Bielski [1], **Szymon Wichary** [2]

[1] Jagiellonian University, [2] Leiden University

Details

Decision-making involving multiple attributes necessitates strategic approaches to effectively process information. Often instead of using rational, complex decision-making strategy, such as the Weighted Additive (WADD) rule, people rely on simple choice heuristics such as Take The Best (TTB). This study investigated how urgency, manipulated as the information presentation rate in a multi-attribute choice task, influences individuals' speed of pre-decisional information processing and the preference towards complex vs. simple decision strategy. We proposed that a sense of urgency, triggered by slower information presentation, would impact the choice of strategy—specifically, slower presentation rates would lead to shorter decision times and reliance on the simple choice heuristics focusing on the most relevant pieces of evidence. Based on previous research, we also proposed that the brain inhibition network would be involved in this process. To explore this hypothesis, we analyzed fMRI data from 24 participants (females=14, age=25.875) using a 3T MRI scanner (TR=1.6s, voxel size=2 mm^3). Participants performed a decision-making task under two conditions: 1) a fast condition with a piece of evidence presented every 1 second, and 2) a slow condition with a piece of evidence presented every 2 seconds. Behavioral data analyzed via a mixed-effects linear model indicated that slower information presentation correlated with shorter response times (coeff=-.021, CI = [-.03, -.011], p-val<.001) . Also, in the slow condition, participants were less likely to use the complex rational WADD strategy (t-stat=2.62, p-val<.01). Furthermore, a GLM-based subtraction analysis highlighted significant activity differences in the frontal and parietal cortices including right inferior frontal gyrus (rIFG) (peak-level: Z-score=5.28, MNI coordinates = (34,28,2)), anterior cingulate cortex (ACC) (peak-level: Z-score=5.05, MNI coordinates

= (8,40,18)), and right supramarginal gyrus (FWE-corrected: Z-score=5.98, MNI coordinates = (42,-44, 44)) in the slow condition. These findings suggest that a heightened sense of urgency is associated with a greater tendency to process pre-decisional information faster, a greater tendency to use simple choice strategy TTB and the greater recruitment of cognitive control brain areas, including the rIFG, postulated earlier as part of the motor inhibition network. Overall, our results demonstrate that urgency profoundly influences both the strategic approach to pre-decisional information processing and the resulting decision-making strategies, underscoring the importance of the external context in decision-making tasks.

P2-I-127 - Computations underpinning the integration of value and perceptual evidence during decision making

Basile Garcia [1], **Valentin Wyart** [2], **Daphne Bavelier** [3]

[1] Université de Genève, [2] Ecole Normale Superieure, [3] Faculty of Psychology and Education Science, University of Geneva

Details

Objectives

Value-based and perceptual decision-making (DM) are ubiquitous in real-life decisions. Value-based DM involves associating subjective values to stimuli, through a trial-and-error learning process, to ultimately select the most rewarding. Perceptual DM involves collecting evidence on stimuli based on perceptual dimensions and then making a decision. Arguably, most decisions involve a tradeoff between these two processes. Yet, these two areas of behavioral research have been largely investigated independently. Most value-based and perceptual tasks share a common structure (typically, a valuation followed by a choice stage), thus it has been argued that one can integrate value-based and perceptual DM into a unified framework. Task designs where such a framework is assumed, though, rarely implement the two processes simultaneously (i.e., a choice that requires perceptual and value-based assessment). As a result, behavioral or computational accounts of how humans weigh value and perceptual information during DM remain scarce. The following study aims at bridging this gap by implementing a gamified task where such naturalistic DM occurs.

Methods

Participants (N=100) played a two-armed bandit task translated as a space shooter game. Options were represented as spaceships that participants had to shoot, with value-learning of each spaceship average rewards. The perceptual cues for their part were possible forcefields in front of spaceships, protecting them from being shot. Forcefields were of different textures, composed of white and black dots, where proportion indicated the probability of destroying the forcefield and thus the spaceship it protected. We modeled participants' behavior as a Bayesian process. A perceptual inference module learns forcefields' probability of being destroyed, while a value-learning module estimates spaceships' average reward. Both estimates are later integrated in decision through two weighting parameters. When $\beta p > \beta v$, the model predicts an overweight of perceptual information. Conversely, when $\beta v > \beta p$, value information is overweighted.

Results
We show that participants display an asymmetric weighting of value and perceptual signals, where perceptual information tends to be prioritized (βp > βv, T=5.29, P=0.00008, d=2.52). Ultimately, we demonstrate that relaxing value-learning complexity allows for more optimal and integrative decisions.

P2-I-128 - The use of graphical visualizations to support decisions: a study to investigate cognitive load.

Maria Clara De Oliveira Gê [1], **Lucia Reis Peixoto Roselli** [2], **Anderson Lucas Carneiro De Lima Da Silva** [1], **Jefferson Sousa Medeiros**, **Laura Emily Ataíde Barros** [1], **Ana Paula Cabral Seixas Costa** [1], **Adiel de Almeida** [1]

[1] Universidade Federal de Pernambuco, [2] Federal University of Pernambuco

Details

The theory of bounded rationality, argues that humans often struggle to evaluate all options when the number of items exceeds the capacity of cognitive processing, as this varies depending on the overall cognitive load. Studies indicate that people may be able to maintain about seven items (plus or minus two) in working memory at once. This underscores the importance of implementing strategies to deal with complexity, such as simplification, segmentation, and the use of decision support tools.

Objective: This study sought to analyze the cognitive load of decision-makers when they have to compare items using graphical and tabular visualizations with multiple attributes and investigates the impact of using graphical and tabular visualizations to support a decision problem of several items disposed in different graphical representations. In addition, two strategies have been considered: choice and elimination decision process.

Methodology: Some experiments have been conducted in the Neuroscience for Information and Decision Laboratory-NSID using a 32-channel EEG and Eye-Tracking Tobii X-120 Eye-Tracker. The sample has been composed of more than 200 undergraduate and graduate students. Through the experiment, participants were asked to solve decision-making problems without specific context. Analyses involving eye tracking include fixation times, heat maps, and pupil dilation. Regarding EEG, Alpha and Theta brain potentials were analyzed.

Results: Statistical tests were performed and significant results were observed. The results indicated that decision-makers present similar performance in compared tables and bar graph. As for gender, women showed better performance and greater cognitive effort. As for the problem, DMs showed a preference for choice, while they performed better on elimination. The problem of elimination also proved to require less effort. Comparing brain responses as for the analyses with Electroencephalogram data, some differences were observed in power within the Alpha and Theta bands concerning the quantity of items evaluated.

Conclusion: This research has the potential to contribute to studies in the field of cognitive neuroscience by providing valuable insights into issues involving assessment and decision-making with multi-attribute, aiming to integrate these findings into decision support systems (Roselli & Almeida, 2022).

P2-I-129 - BEHAVIORAL STUDY OF THE PREFERENCES EXPRESSED BY TWO DIFFERENT GROUPS OF DECISIONS-MAKERS.

Lucia Reis Peixoto Roselli [1], **Jefferson Sousa Medeiros**, **Maria Clara De Oliveira Gê** [1], **Rodrigo José Leite Cavalcante** [1], **Antônio Reinaldo Silva Neto** [1], **Thárcylla Rebecca Negreiros Clemente** [1], **Anderson Lucas Carneiro De Lima Da Silva** [1], **Ana Paula Cabral Seixas Costa** [1], **Danielle Morais** [1], **Adiel de Almeida** [1]

[1] Universidade Federal de Pernambuco

Details

Glimcher (2014) and Eagleman (2015) argue that human decisions are influenced by various elements, such as past experiences and environmental stimuli. Understanding how factors influence behaviors and decision-making is fundamental to improving decision-making processes.

Objective: This study aims to investigate decision-makers' (DM) behavior in a multi-attribute decision-making problems by examining how preferences are expressed and their relations with the degree of knowledge in the area. This approach enables a deeper understanding of decision process, seeking valuable insights to advise DMs and contributing to the enhancement of Decision Support Systems (DSSs).

Method: To investigate participants' behavior and attention during decision processes, the X-120 Eye-Tracking has been used. In the study, the sample has been composed of two groups of participants. The group of control considered individuals with knowledge in multi-attribute approach, and the other group considered participants without previous knowledge, totaling 42 participants. The experiment involved the collection of preferences expressed by decision-makers in order to solve their own problems.

Results: Wilcoxon and Mann-Whitney statistical tests were carried out, revealing different patterns in eye movements and pupil size between the groups with and without knowledge for different levels of significance (0.1, 0.05 and 0.01). The group of control presented analytical and focused behavior, especially in tasks related to analyzing attributes and interpreting graphs, suggesting that they perform these tasks with higher cognitive effort The group without knowledge, on the other hand, showed greater attention, cognitive effort and more intense fixation in the initial stages indicating that they try to understand the decision process. In addition, over time the fixations decreased in the area of graphs, suggesting that they were probably trying to reduce the effort put into the process by focusing on the choice they had to make.

Conclusion: The results emphasize the importance of customizing information presentation according to the user's knowledge, thus facilitating data understanding. It is recommended that for participants without knowledge more instructions should be made at the begging of the process. For the DSS, it is suggested to include informative screens before each stage, with brief descriptions of what will be done

next. It is also beneficial to add a help button, offering a more detailed step-by-step guide for decision-makers in case of doubts during the execution of the decision-making process.

P2-I-130 - Intrinsic and Instrumental value: Domain-general principles determine choice across material, cognitive and visual domains.

India Pinhorn [1], **Bastien Blain** [2], **Tali Sharot** [1]

[1] University College London, [2] Panthéon-Sorbonne University

Details

Objective: What makes something rewarding? Do humans use the same principles and computations to determine how rewarding new shoes are, as they do to determine how rewarding reading a Shakespeare sonnet is, or viewing Monet's Water Lilies? Here, we test the hypothesis that the brain uses domain-general principles to calculate the value of 'things', regardless of whether those 'things' are material (e.g., an eraser), cognitive (e.g., a piece of trivia), visual (e.g., a painting).
Methods: Participants (Exp 1 N = 59; Exp 2 N = 115) rated material, visual and cognitive items on different 'reward' dimensions, derived from the psychological literature. We collected two canonical behavioral signatures of reward – liking and wanting. Participants rated how much they liked each item and selected between items. We performed a dimensionality reduction (PCA) within each domain (Visual, Cognitive, Material), which allowed us to assess domain-generality in the structure of reward dimensions. In Mixed-Effects models within each domain, the resulting components were then used to predict liking and choice.
Results: Dimensionality reduction revealed that across domains (Visual, Cognitive, Material) the value of an item was described by the same two components: one which describes an item's instrumental utility (i.e., how 'useful' it is) and one that describes its 'intrinsic value' (which encompasses beauty, interest and meaning). In each domain, these 'Intrinsic' and 'Instrumental' components explained the most variance (Exp 2: Intrinsic: 40% (Visual), 32% (Cognitive), 40% (Material); Instrumental: 24% (Visual), 31% (Cognitive), 21% (Material)). Both components predicted how much participants liked an item (Exp 2: Intrinsic (Visual: $\beta = 0.78 \pm 0.07$, $t(79.97) = 44.91$ $p < 0.001$; Cognitive: $\beta = 0.52 \pm 0.08$, $t(94.25) = 28.21$, $p < 0.001$; Material: $\beta = 0.60 \pm 0.09$, $t(43.6) = 23.12$, $p < 0.001$); Instrumental (Visual: $\beta = 0.20 \pm 0.07$, $t(92.08) = 9.91$, $p < 0.001$; Cognitive: $\beta = 0.56 \pm 0.08$, $t(105.71) = 28.40$, $p < 0.001$; Material: $\beta = 0.28 \pm 0.09$, $t(61.3) = 12.62$, $p < 0.001$)) and whether they chose it (Exp 2: Intrinsic (Visual: OR = 5.88 ± 2.30, $p < 0.001$; Cognitive: OR = 3.13 ± 1.28, $p < 0.001$; Material: OR = 3.32 ± 1.09, $p < 0.001$); Instrumental (Visual: OR = 1.47 ± 0.35, $p < 0.001$; Cognitive: OR = 3.87 ± 1.27, $p < 0.001$; Material: OR = 2.22 ± 0.71, $p < 0.001$)).
Conclusion: The array of 'things' humans value is vast, and it is unclear how the brain distills complex reward attributes into representations that facilitate efficient decision-making. We show that intrinsic and instrumental value may provide domain-general principles to organize such representations.

P2-I-131 - Optimistic Update Bias Beyond Life Events

Zheyi Ding [1], **Jian Li** [1]

[1] Peking University

Details

Objective: This unrealistic optimism in belief updating, that people update beliefs more in response to good news than bad news, had been replicated in many studies adopting belief updating task focused on adverse life events. However, debates existed over the robustness of these findings after considering positive or neutral life events, arguing that the observed optimism bias may be caused by methodologies flaws. In light of this controversy, we used monetary stakes instead of life events, to ensure balanced experimental settings across different conditions and investigate belief updating in a laboratory-controlled manner.

Methods: We created different scenarios to let participants estimate the probability of gains and losses based on provided evidence. In experiment 1 (N = 50), participants estimated probabilities after observing dot stimuli presented in low and high resolutions separately. Experiment 2 (N = 50) was similar to experiment 1, except that participants made two estimates after viewing partial and complete stimuli. Experiment 3 (N = 51) introduced a scenario where participants estimated the likelihood from jars they owned, with second estimate made after knowing the base rate. Unknown to participants, the second screen of information was manipulated based on participants' first estimate. We focused on participants' belief updating in response to different valence of information, defined as the difference between the first and second estimate. 2 (gain, loss) by 2 (good news, bad news) repeated-measures ANOVA and linear mixed effect models were used for main analysis.

Results: Experiment 1 and Experiment 2 showed that participants updated their beliefs symmetrically between different conditions. The updates did not differ between the good news and bad news condition, after controlling the baseline (Experiment 1: $F(1,49) = 0.45$, $p = 0.506$, $\eta p2 = 0.009$; Experiment 2: $F(1,49) = 0.07$,$p = 0.792$, $\eta p2 = 0.001$). In Experiment 3, participants showed asymmetric belief updating. There was a significant main effect of information desirability ($F(1,50) = 5.18$, $p = 0.027$, $\eta p2 = 0.094$). Linear mixed models produced similar results. There was a main effect of information desirability, with participants updating more in the face of good news than bad news ($\beta = 0.81$, $SE = 0.17$, $t = 4.78$, $p < 0.001$).

Conclusion: The results of our current study demonstrate the validity of optimism bias and that the optimistic update bias can be generalized beyond real-life events. Meanwhile, the optimism bias is absent in the perceptual observing paradigm, which indicates the necessity of sense of ownership in this cognitive bias.

PS.02.02, P2-I-132 - Creative choices rely on subjective values and suffer from cognitive biases

Sarah Moreno-Rodriguez [1], **Alizée Lopez-Persem** [1], **Emmanuelle VOLLE** [2]

[1] Paris Brain Institute, [2] Sorbonne University

Details

Creativity is defined as the ability to produce ideas that are both original and adequate. According to the dual-process theory, creativity involves a generation phase, during which individuals produce candidate ideas, and an evaluation phase, during which individuals assess and select the best idea among the candidates. Previous research (Moreno-Rodriguez et al., *in prep*) indicates that this evaluation phase features the subjective valuation of candidate ideas via the human reward system, also known as the Brain Valuation System (BVS). The current study's goal was to further investigate the decision-making mechanisms involved in creative thinking by focusing on the selection component of idea evaluation and to explore the occurrence of potential cognitive biases, such as an undue preference for adequacy over originality in creative ideas.

Alongside assessments of creative abilities and performance, 111 participants completed an idea production task, a subjective value rating task, a binary choice task and a second rating task, this time regarding the ideas' originality and adequacy level. 40 participants of our sample underwent this experiment in an MRI scanner.

At the behavioral level, we used computational modeling to explain individual ratings using the following utility function:

$$\text{subjective value}_{idea} = (\text{alpha} \times \text{Originality rating}_{idea}^{\text{delta}} + (1-\text{alpha}) \times \text{Adequacy rating}_{idea}^{\text{delta}})^{1/\text{delta}}$$

This model revealed that individuals followed different valuation patterns depending on the task: in their subjective value ratings, individuals integrated originality and adequacy in a balanced manner (alpha~0.5); but in their choices, individuals unreasonably preferred adequacy at the expense of originality (alpha<0.5).

Interestingly, the size of this bias varied among individuals, and those with higher biases against originality scored lower on creativity scores than those with lower biases (correlation between bias and creativity score: r=0.52, p=0.025), highlighting the interplay between valuation patterns and creative abilities.

These findings prompt consideration of which valuation patterns prevail in creative thinking: when producing ideas, do individuals follow their declarative, unbiased preferences (as in their ratings) or their implicit, biased preferences (as in their choices)? Current analyses including model comparisons of BVS signal during the idea production task are addressing this question.

Overall, this study challenges current accounts of the neurocognitive bases of decision-making during creativity and underscores the importance of employing neuroeconomics methods to explore creativity and its cognitive biases.

P2-I-133 - Developing and validating a scenario-based cognitive bias questionnaire

Chengrui Zhou [1], **Yi Huang** [1], **Rongjun Yu** [2]

[1] Lingnan University, [2] Hong Kong Baptist University

Details

Introduction and Objectives: Human decision-making often suffers from cognitive biases, leading to suboptimal outcomes across various domains. Existing tools to evaluate biases lack comprehensiveness and reliability, highlighting a crucial gap in research. To address this, we aim to develop a comprehensive questionnaire covering diverse biases in decision-making, leveraging advanced deep learning methods for item clustering and selection.

Methods: Our methodology includes an in-depth literature review, content validity assessment through expert input, pilot testing, and data analysis using exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and item response theory (IRT). We identified 230 biases, refined to 92 items using deep learning clustering techniques, and conducted EFA to reveal five factors explaining 75% variance.

Results: The developed scale exhibited strong reliability (Cronbach's alpha 0.909) and validity (CFI > 0.9) with pilot studies (N=87) across the five factors: Belief, Social, Memory, Information, and Calculation biases. Item response theory (IRT) analysis identified and removed 12 low-discriminating items, enhancing the scale's efficacy in measuring cognitive biases.

Conclusion: Our study addresses critical gaps in existing tools for measuring cognitive biases, providing a robust and comprehensive instrument developed using advanced deep learning techniques. By advancing decision science and highlighting individual differences in cognitive bias, this research contributes significantly to improving decision-making outcomes.

P2-I-134 - Dissociating the role of the peri-hippocampal and parietal cortex in human model-based reinforcement learning

Fabien Cerrotti [1], **Alexandre Salvador** [2], **Sabrine HAMROUN** [1], **Maël Lebreton** [3], **Stefano Palminteri** [1]

[1] École Normale Supérieure, [2] Ecole Normale Supérieur de Paris, [3] Paris School of Economics

Details

Introduction

Many real-life decisions involve a tension between short-term and long-term outcomes. In the reinforcement learning framework, an optimal solution to this trade-off is to compute prospective values using a model of the environment (model-based RL). Experimentally, this trade-off is often instantiated by multi-step decision-making problems, where decisions taken at an initial stage concomitantly generate an outcome and the transition to a new state. The present study aimed to compare the neural correlates of choices involving a prospective value computation. To this end, a new version of a two-step learning task was administered to participants (N=28) undergoing fMRI scanning.

Method

Our two-stage Markov decision task was inspired by previous studies (Daw et al., 2011; Gläscher et al., 2010). On each trial, participants began in one of two possible initial states. There, they were presented

with two options marked by abstract symbols from the Agathodaimon alphabet. Participants had to select one option, each of which was associated with a different probability of leading to one of two possible final states. One of the two final states was in average more rewarding than the other (the "rich state"). States were denoted by different colors. In the final step, participants had to choose between two new symbols, one being more often rewarding than the other.

Importantly, the probability of transitioning to the rich final state was higher when the less rewarding option was chosen in the entry state (leading to a trade-off between short- and long-term rewards).

Results

Behavioral analyses (N=57) revealed that participants had learned both the reward values associated with each state ($X^2(57) = 10.044$, $p<.001$) and an explicit model of the task, namely the state-transition probabilities, $W(57)=1058$, $p<.001$.

A contrast was made between the neural activity of the two entry steps and the two final steps at the moment of choice and outcome. At the moment of choice, the peri-hippocampal cortex demonstrated heightened brain activity during the first step in comparison to the rich state ($t(25)=5.97$, $p<.001$) and the poor state in the second step ($t(25)=6.49$, $p<.001$). At the moment of outcome, the parietal cortex demonstrated increased brain activity between the first step and the rich state ($t(25)=9.11$, $p<.001$) and the poor state ($t(25)=6.07$, $p<.001$) in the second step.

Conclusion

The results provide new insights into the neural underpinnings of model-based reinforcement learning, suggesting a specific role for the peri-hippocampal cortex in forward planning and the parietal cortex in choices involving immediate rewards.

P2-I-135 - Disambiguating the effect of individual risk preference on incentivize-induced confidence bias

Mandy Renfro [1], **Cassandra Raymond** [1], **Corey Plate** [1], **Zoe Boundy-Singer** [2], **Corey Ziemba** [2], **Silvia Lopez-Guzman** [3]

[1] National Institute of Mental Health, [2] University of Texas at Austin, [3] National Institutes of Health

Details

Objective. Incentivization is an effective way to influence immediate choice behavior in both human and animal subjects. Previous research has shown the introduction of reward incentives can bias reports of decision confidence. If one considers incentivized confidence reporting as a lottery, high confidence can sometimes be *risky*. Here, in two studies, we sought to determine how an individual's monetary risk preference may moderate the effect of incentives on reports of perceptual confidence, and whether confidence behavior is more consistent with expected versus subjective value maximization strategies.

Methods. In Study 1, participants (n = 149) recruited from CloudResearch completed two online computerized decision-making tasks which assessed: (1) perceptual orientation discrimination, and (2) risk preference. The *perceptual task* consisted of two 200-trial incentive blocks (i.e., low and high) in which participants indicated the orientation of a visual pattern (i.e., right or left) and their degree of confidence (i.e., low or high). In high (low) incentive blocks, payout for high confidence correct responses was $13 ($9), while all other payouts were identical across blocks. In Study 2 (n = 33), we introduced a control block with no incentives. To ensure incentive compatibility across trials/tasks, a single trial from one of the tasks was randomly selected for payout. Risk aversion was estimated for each participant by fitting a power expected utility model. Confidence reports were fit to the CASANDRE process model of confidence, which returned individual parameters for confidence criterion and metacognitive precision.

Results. Participants reported higher confidence in high- versus low-incentive blocks (mean difference in criterion = 0.39, $p < 0.001$). While the relationship between confidence bias and risk preference was not significant (r = 0.07) for Study 1, in Study 2 we found participants appeared to shift from maximizing expected value in low incentive blocks to maximizing subjective value in high incentive blocks.

Conclusions. Our study highlights the interplay between confidence and individual differences in risk preference. Further, the ability to disambiguate risk preference from confidence may prove useful in improving our understanding of behavior in clinical populations where extreme risk preferences or abnormalities in confidence express, such as in mood, anxiety, or substance use disorders.

P2-I-136 - Cognitive imprecision explains choice stochasticity across different preference tasks

Clara Haeffner [1], **Miguel Barretto García** [2], **Silvia Lopez-Guzman** [3]

[1] National Institute of Mental Health, [2] Washington University in St. Louis, [3] National Institutes of Health

Details

Objective: Choice is stochastic, but the extent to which choice stochasticity is conserved throughout different types of decisions—and thus alludes to a common neural mechanism and a trait-like nature of decision noise—is unknown. Current computational models of value-based choice model choice probability as arising from a linear value comparison that is passed through a stochastic choice function. Alternatively, recent work in cognitive imprecision (CI) suggests that noise arises from neurocognitive representations of magnitude. CI suggests a common neurocognitive mechanism of imprecision, which can link decision noise to risk attitudes. Here, we test whether CI can plausibly account for various levels of choice stochasticity across tasks. We evaluate the intra-individual stability of choice stochasticity across a risk-attitude (RA) and delay discounting (DD) task, and interrogate the role of CI in accounting for this relationship. **Methods:** 42 healthy adults participated (ongoing analyses, n=460). In the RA task, participants selected between a guaranteed $5 gain or a lottery where they could win a different amount. In the DD task they chose between a smaller amount of money at a short delay, and a larger amount of money at a longer delay. The values, delays and probabilities in each task varied trial-by-trial. We fit the RA data using an expected utility (power) model while the DD data was fit to a hyperbolic discounted utility model. Choice probability was modeled as a linear value difference (VD)

softmax model and a log-linear CI model. The VD approach assumes a perfect value comparison injected with noise during the choice process, while CI assumes that noise is positively scaled with value magnitude. We used maximum likelihood estimation and hierarchical Bayesian modeling and performed Spearman rank correlations for statistical analyses. **Results:** Choice stochasticity varies across individuals irrespective of the modeling: In VD, RA stochasticity ranged from 0-8.0 and DD stochasticity from 0.1-8.0. The CI model showed ranges of 0.1-0.9 for RA stochasticity and 0.01-1.3 for DD stochasticity. Crucially, we found that regardless of framework, stochasticity correlated across tasks (VD: $r(40)=0.65$, $p<0.001$; CI: $r(40)=0.31$, $p<0.05$). VD does not offer an a priori assumption as to the nature between stochasticity, risk, and discounting, but CI does. CI offers precise predictions on the relationship between stochasticity, risk, and discounting across individuals. Overall, cognitive imprecision may serve as a general mechanism of choice stochasticity that could account for individual risk attitudes and discounting behaviors.

P2-J-137 - Application of Neuroimaging to Better Understand the Role of Emotions in Patient and Caregivers' Decision-Making

Eline Van Den Broek-Altenburg [1]

[1] University of Vermont

Details

Objective

This study uses functional magnetic resonance imaging (fMRI) in health economic models to better understand the role of emotions in patient decision-making. The large-scale sharing of task-based functional neuroimaging data has the potential to allow novel insights into behavioral mapping for the use of health economic evaluation. The objective is to provide "proof of concept" of the added value of fMRI data in health economic evaluation among patients with Alzheimer's and Alzheimer Related Dementias and their caretakers as a scalable solution to improve access and lower costs.

Methods and Results

Our recent work on patient preferences, using discrete choice experiments (DCE), has established an important role of emotions in health treatment preferences. We propose to use fMRI to better understand how emotions contribute to individual, and societal preferences in health with a case study in a cohort of 10 patients suffer from Alzheimer's Disease and Alzheimer's Related Dementias (AD/ARD) and 20 of their caregivers. We will collect states preferences data through a DCE to analyze trade-offs patients make when choosing between treatment or policy options where these hard to measure attributes are important. A DCE provides the opportunity to estimate pair-wise choices and analyze marginal values or the total value of a health service or good. The survey will also ask questions about emotional state at the time of choosing. SP data from the DCE will be augmented with study data from the task-based fMRI. We will use fMRIPrep to prepare the data for statistical analysis.

For our analytic approach, we will use a mixed logit model to predict take-up rates as a function of attributes of the choice and attributes from the DCE data. We will calibrate the model to actual

psychometrics values from the population we derived from the survey. We will use random effects logit models to detect changes in emotion and changes in preferences for AD/ARD treatment options.

Conclusions

We expect to better target treatment decisions to individual patients' or caregivers' needs and preferences and show through renewed cost-effectiveness analysis, including this data, that higher value care can be delivered for lower costs. If successful, using functional neuroimaging and connecting to psychometric and behavioral data might become the future standard for health economic evaluation focused at improving access and affordability in health.

P2-L-138 - Excessive-demand measure outperforms in explaining lab asset-market price changes: Toward a biomarker of excessive demand

John Haracz [1]

[1] Indiana University

Details

Objective: Dynamic stochastic general equilibrium models have been criticized for failing to forecast the Global Financial Crisis (Yellen, 2010; Vines & Wills, 2020). This and other flaws of neoclassical economics were proposed to arise partly from the failure of equilibrium-based models to capture *excessive demand* (Haracz, 2023), which exceeds the balanced "excess demand" in general equilibrium theory (Arrow, 1974, p. 266; Debreu, 1984, p. 270). Excessive demand is defined as demand that promotes disequilibria in asset or goods markets and drives prices above fundamental values (e.g., an asset-price bubble). Neuroimaging studies are elucidating the neuroeconomics of asset-price bubbles (Smith, A. et al., 2014). However, these studies have been limited in characterizing individual-level brain-behavior relationships due to the lack of a subject-level excessive-demand measure (EDM). The present study makes such a measure available to neuroimaging researchers. Methods: In a standard lab asset-market design (Smith, V.L. et al., 1988), 9 experiments each included 9 undergraduate-student subjects. Experiments consisted of 15 2.5-minute periods of trading an asset with a fundamental value that declined across periods. To capture excessive demand, the end of each Period 1-14 was followed by a survey that elicited each subject's number of asset shares that they want to hold at the end of the next period. This measure was designed to tap into anticipatory affect that may drive price bubbles. Two other predictive measures included excess bids (Smith et al., 1988) and momentum, which was measured as the most recent inter-period price change. Results: The EDM, which explained 34.5% of the variance in asset-price changes, significantly outperformed the excess-bids and momentum measures, which each explained less than 10% of this variance. Conclusions: The EDM's outperformance in predicting price changes aligns with numerous other findings that underscore the predictive power of measures related to anticipatory affect. For example, fMRI-measured activity in nucleus accumbens, an area implicated in anticipatory affect, performed better than choice behavior in forecasting crowdfunding outcomes (Genevsky et al., 2017). Similarly, the survey-elicited EDM, which may reflect anticipatory affect, was a better price-change predictor than the behavioral excess-bids measure. Therefore, the presently introduced EDM may facilitate finding an excessive-demand biomarker with market-level predictive power. Biomarkers for neuroforecasting market-level outcomes in luxury-goods

markets (e.g., top-line cars, flights, etc.) may elucidate the neuroeconomics of global warming (Haracz, 2024).

Poster Session III

P3-A-140 - Economic oracles: how markets discover optimal solutions to complex problems

Hassan Andrabi [1], **Robert Woods** [2], **Konstantinos Ioannidis** [2], **Peter Bossaerts** [2], **Nitin Yadav** [1]

[1] University of Melbourne, [2] University of Cambridge

Details

Objectives. When faced with prohibitively difficult problems, humans attempt to cast the original problems to simpler reductions. Prior work suggests that this occurs through the use of approximation algorithms, that constrain the search to only a small subset of the total solution space. While humans individually search among these small subsets, collectively they visit a large fraction of the solution space. This naturally invites research into how humans could communicate, most effectively and reliably, in order to share knowledge. Guided by the notion of oracles in computer science, we study here to what extent markets can fulfil this role to spread knowledge about the solutions to computationally hard problems. **Methods**. We performed a series of experiments in which participants (N = 99) attempted to solve instances of the Knapsack Problem (KP), a canonical example of a complex task, while sharing information about their solutions by trading in an experimental market. In each session, we presented all participants with the same instance of the KP, and allowed them to trade units of an experimental asset that promised a payoff equal to the best solution discovered among all participant in that session. In this way, participants could earn money by purchasing the experimental asset at prices below the best solution found, and by selling it at prices above the best solution. **Results.** We find overwhelming evidence that the design works, not only in spreading the best solution attained among participants, but also in revealing the true optimal solution. In all experiments, a significant fraction of participants (usually more than 2/3) discover the optimal solution, so knowledge spreads widely. This is true even for instances with 24 items, in which the true optimal solution is just one out of almost 17 million possible alternatives. However, speed of convergence deteriorates regardless of number of items when difficulty increases in a metric that predicts individual humans' ability to find the solution. **Conclusions**. Our findings provide ideas on how to incentivise groups of people to try their best at solving hard problems without having to know what the optimal solution is, and in such a way that enables individuals to benefit from the collective knowledge of others. The market solution could be applied, for example, to incentivise fund managers to find the best investment opportunities over a given time horizon.

P3-B-141 - Uncommon Errors: Adaptive Intuitions in High-Quality Media Environments Increase Susceptibility to Misinformation

Reed Orchinik [1], **Cameron Martel** [1], **David Rand** [1], **Rahul Bhui** [1]

[1] Massachusetts Institute of Technology

Details

OBJECTIVE: Digital media environments are often argued to foster belief in misinformation by subverting the process of deliberation and promoting reliance on intuition. A key presumption of this account is that intuition is inherently gullible, based on prominent psychological theories suggesting an immutable default tendency to initially believe content is true that may be subsequently suppressed by deliberation. We propose that this apparent irrationality may result from ecologically rational adaptation to the information environment. When most encountered content is true, one's intuitions should initially favor belief pending the outcome of deliberation.

METHODS: We experimentally tested whether intuitions of truth adapt to the veracity base rate—the proportion of true and false content in an environment—in a simulated social media "news feed" with real headlines. Participants (N = 2,810) were exposed to a sequence of 50 headlines that were either mostly (80%) true or mostly false in an initial "habituation" phase, followed by an evenly-split true/false "evaluation" feed for 30 headlines (without being notified of the change). They indicated the perceived accuracy of each headline.

RESULTS: Consistent with the use of veracity base rates, participants demonstrated differences in errors. Those exposed to mostly true headlines were 15% or 5.6pp ($t=11.8$, $p<0.001$) more likely to misidentify false headlines compared to true ones, and vice versa for those seeing mostly false headlines (5.5% or 2.8pp, $t=4.4$, $p<0.001$). A hierarchical Bayesian drift-diffusion model decomposed these effects cognitively. There were no observable differences between conditions in the non-decision time ($D=0.02$, $p=0.92$; two-sample KS test) or boundary ($D=0.03$, $p=0.65$) and a slight difference in the drift rate ($D=0.06$, $p=0.02$). However, the bias was significantly greater in the mostly true condition ($D=0.30$, $p<0.001$). Fitting DDMs on data split into bins of 10 rounds revealed a clear divergence of the bias across the habituation phase. Some convergence was seen after the switch to the 50/50 evaluation phase.

CONCLUSION: Our results support the ecological account of intuitive belief in misinformation and contrast with the traditional theory of a fixed intuitive bias. This work addresses how intuitions about credibility are formed, how this plays a role in the process of deliberation as reflected in models increasingly used in neuroeconomics, and how intuition and deliberation contribute to an important low-attention naturalistic judgment.

P3-B-142 - I want that!: Investigating the EEG Signals Underlying and Preferential Choice during Purchase Decisions

Youngbin Kwak [1], **Francesca Walsh** [1]

[1] University of Massachusetts Amherst

Details

Objective: Amplitude of classic EEG motor signals such as the readiness potential have shown to encode decision variables such as reward and probability during economic decision making. Not many studies have investigated how such effect translate to real-world purchase decisions. The current study aimed

to investigate how preference is encoded in the temporal cascade of EEG neural signals across the motor hierarchy.

Methods: We used a novel, two-choice "shopping" paradigm to investigate the role of central, parietal and lateralized motor signals during a preferential two-choice task (n = 43). During the task, an item and a monetary amount was sequentially presented on the center of the screen, with the order of presentation randomized across trials. Since the monetary value is fixed across trials, one's purchase decision is made at the time of item presentation. A motor cue followed upon which left or right hand button press was made to choose between money and item. There were two block types: 1) where the left-right mapping to item-money was fixed across trials were fixed, such that advance motor preparation was allowed before the motor cue (*preparation*), and 2) where the hand-choice mapping changed, such that advance motor preparation was not possible before the motor cue (*no preparation*). We compared the neural signals while processing the items across the chosen vs. unchosen, preparation vs. no preparation and during item presentation vs. money presentation.

Neural Results: EEG signals from the central electrode reflecting activity from secondary motor area (e.g., pre-SMA and SMA) predicted whether or not the item will be chosen specifically at the time when choice decision could be made, but not before. Such pattern was more significant when advance preparation could be made than when no preparation could be made. Similar results were found in the parietal electrode.

Discussion:

These results provide evidence suggesting the neural signals from central and parietal electrode can reliably predict choice while participants are still evaluating the options. This effect was specifically present when advance motor preparation was possible. Further analyses will be performed in the lateralized motor signal reflecting primary motor area to investigate the encoding of preference across the motor hierarchy.

P3-B-143 - Color Overload: Exploring the Effects of Design Complexity on Sustainable Consumption

Martina Columbano [1], **Cristina Zerbini** [1], **Donata Tania Vergura** [1], **Beatrice Luceri** [1]

[1] University of Parma

Details

Packaging design significantly influences consumer decision-making, but when visual elements like color numerosity and design complexity are excessive, could they cause cognitive overload? This occurs when the information presented surpasses consumers' cognitive capacity thus leading to impaired information processing. The present project seeks to investigate the impact of color numerosity and design complexity on consumer perceptions of sustainability, healthiness and tastiness within the framework of cognitive overload. By understanding how these visual elements on the packaging contribute to or mitigate cognitive overload, we aim to provide insights into optimizing packaging design that balances aesthetic appeal with functional clarity, thereby enhancing consumer satisfaction and promoting sustainable consumption.

The proposed research will be conducted in three stages. The first stage comprises of two experiments Study1 and Study2. Study1, will focus on color numerosity (low-medium-high), with a between subject three cell design experiment (n=200), to assess its effect on the dependent variables (purchase intention, perceived sustainability of both the product and the packaging, perceived healthiness, perceived tastiness). Study2, will evaluate the impact of design complexity (low-medium-high) on the dependent variables in a between subject three cell design (n=200). The second stage, Study3, will involve a 3x3 between subject design experiment incorporating both color numerosity and design complexity (n=500). Lastly, in the final stage (Study4) eye-tracking metrics will be incorporated in the design (n=225) to understand how consumers process these visual elements in real time, in addition to the behavioral variables explored in studies 1-3, Study4 will incorporate metrics such as fixation duration, time to first fixation, visualization patterns and pupillary responses.

The anticipated results are expected to demonstrate that a more complex design (e.g. more colors) could generate a higher cognitive load and, therefore, decreasing sustainability perception. Moreover, the eye-tracking analysis is expected to reveal distinct visual attention patterns based on these two core variables offering insight into how consumers engage with packaging designs as well as difference in arousal between the conditions.

Understanding how color quantity and design complexity affect consumer perceptions is essential for creating packaging that highlights the sustainability-related product attributes without losing appeal. This research will offer valuable insights for marketers and designers, helping them balance aesthetics with environmental responsibility.

Funding provided by PNRR-M4C2- I1.1 – MUR Call for proposals n.104 of 02-02-2022 - PRIN 2022 - ERC sector SH1- Project title: Sustainable Food Labels for Product Communication - Project Code 2022BWNJBA - CUP Code D53D23006390006 - Funded by the European Union – NextGenerationEU.

P3-B-144 - Neuroforecasting changes in aggregate demand for vehicles

Dylan Christiano [1], **Cynthia Wu** [2], **Tara Srirangarajan** [2], **Shabnam Hakimi** [3], **Matthew Klenk** [3], **Charlene C. Wu** [3], **Brian Knutson** [2]

[1] University of Michigan, [2] Stanford University, [3] Toyota Research Institute

Details

Objective: Demand for vehicles using alternative fuel sources is increasing, leading to a proliferation of new vehicle models and types. Little is known, however, about how market demand might change for these vehicles. In two studies, we sought to examine whether brain activity could predict individuals' desire to purchase and learn more about vehicles (Erk et al., 2002), as well as forecast changes in demand for vehicles out-of-sample in the US (Knutson & Genevsky, 2018).

Method: Two cohorts of subjects interested in purchasing a vehicle in the next three years were scanned at the same time point, May/June, in two consecutive years, 2022 (n=13) and 2023 (n=42). Subjects participated in a vehicle rating task as their brain activity was monitored using Functional Magnetic Resonance Imaging (FMRI). During task trials, subjects first saw a centrally presented image of the vehicle and its' name (2 sec), followed by its fuel source (2 sec), followed by scales on which they could indicate their interest to learn more about and desire to purchase (4 sec each). Individual analyses sought to predict rated desire to know more about each vehicle based on Volume of Interest data from the Nucleus Accumbens (NAcc), Medial PreFrontal Cortex (MPFC), and the Anterior Insula (AIns) (Samanez-Larkin and Knutson, 2015). Aggregate analyses sought to use VOI data averaged over the sample to forecast change in demand for units sold over 2022 and 2023 (indexed with data from goodcarbadcar.com).

Results: Individual choice and whole brain analyses revealed that subjects' initial NAcc response to vehicles predicted wanting to know more about and purchase those vehicles. Activity in other volumes of interest did not, however, predict wanting to know more about or purchase vehicles. Average NAcc response to vehicles forecast the slope in market demand of units sold in both years (2022 $p<.006$; 2023 $p < .0475$), but not average units sold in either year. Average desire to know and purchase ratings, however, did not forecast the slope or average units sold. In a regression model combining behavioral ratings with NAcc activity with behavioral ratings, only NAcc activity significantly forecast the slope of units sold in both years.

Conclusions: Early neural responses to vehicles in the NAcc predicted individuals' desire to purchase and learn more about them. Further, average group NAcc activity replicably forecast changes in demand for vehicles. These findings suggest that brain activity adds value to conventional measures for forecasting consumer demand for new vehicles.

P3-B-145 - Characterizing Evidence Accumulation Across Simple Choice Domains with EEG

Kianté Fernandez [1], **Michael Nunez** [2], **Ian Krajbich** [1]

[1] University of California, Los Angeles, [2] University of Amsterdam

Details

Background: Economic and perceptual decisions vary along at least two dimensions: subjective versus objective criteria, and evaluation of a stimulus versus a representation. Mathematical models of decision-making suggest that each choice type involves neural computations that accumulate evidence for the choice alternatives. However, the overlap or distinctions between the processes underlying these categories remain largely unexplored. In this study, we examine the decision-making process across all four categories using measures of choice, response times, eye-tracking, and EEG.

Methods: Participants (N=40) were given four tasks: choosing between two foods they preferred to eat, choosing between two food images they preferred, determining which of two foods took up a greater proportion of the screen, and assessing which of two foods weighed more. Each task used the same motor response and set of food images as stimuli.

Results: In line with previous research, we observed consistent behavioral patterns across all categories: participants generally chose in line with their reported ratings, were faster for easier decisions, and chose the option they looked at more. Additionally, we found preliminary evidence of the presence of the centro-parietal positivity (CPP), a late event-related EEG potential that reflects the accumulation of evidence during decision-making. Specifically, we found that the CPP increased with the quality of stimulus evidence, decreased with faster response times, and was higher for correct choices compared to incorrect ones, with the strongest effects observed in decisions involving determining which of two foods took up a greater proportion of the screen. The analysis is ongoing.

Conclusions: Our findings provide model-free evidence for the presence of CPP across choice categories, with decisions involving objective criteria and the evaluation of a stimulus showing the largest effects. With evidence for the CPP being variable (e.g., Frömer et al., 2024), future analysis aims to use model-based approaches to account for potential confounds related to perceptual and decision-related processes overlapping using methods like deconvolution techniques, hidden multivariable pattern analysis, and neurocognitive modeling techniques.

P3-B-146 - Eyes on the Story, Hands on the Wallet?

Andrea Bazzani [1], **Silvio Ravaioli** [2], **Alice Nobilio** [1], **Paolo Frumento** [3], **Ugo Faraguna** [3], **Giuseppe Turchetti** [1]

[1] Sant'Anna School of Advanced Studies, [2] Cornerstone Research, [3] University of Pisa

Details

Eyes on the Story, Hands on the Wallet?

Study's Objective: The aim of this research is to explore the impact of storytelling in e-commerce on consumer behavior using eye-tracking. The study investigates how storytelling affects consumer behavior on online retail websites, considering the balance between minimal information that can enhance focus and excessive details that may lead to information overload.

Methods: An experimental design involved six fictitious online retail websites, three for beer and three for cheese products, with variations in the presence of storytelling on the product and cart pages. A total of 40 young adults (average age 23.72 ± 2.46 years, 12 females) participated in the study, navigating through these websites (no time limit) while their gaze was recorded by an eye-tracker device. Participants' attention to various visual elements (title, picture, technical info, description, storytelling, price) was measured, along with their purchase intentions. Logistic regression models estimated the effect of storytelling and the other visual elements on the likelihood of purchase, adjusting for age, sex, and individual consumption habits for the two product categories.

Results: The findings revealed a greater probability of purchase with increased attention to storytelling (β = 0.084 for beer, β = 0.090 for cheese, p < 0.001 for both). However, storytelling's effectiveness was linked to the amount of attention it received. Specifically, storytelling elements, when focused on for more than 5 seconds, increased the likelihood of purchase, while limited attention (less than 5 seconds) reduced purchase likelihood compared to the no storytelling condition. Among other visual elements, fixation time on price significantly increased purchases in both product categories (β = 1.448 for beer, β = 0.753 for cheese, p < 0.001 for both), while time spent on technical information and on the picture significantly increased purchases only for beer (β = 0.051, p < 0.001; β = 0.126, p = 0.031, respectively).

Conclusions: The results suggest that storytelling, along with price, can be more effective than other visual elements in capturing consumer attention and influencing purchase decisions, provided it receives adequate focus. These findings highlight the need for online retailers to design their websites and product information in ways that optimize consumer attention to storytelling elements. The use of consumer neuroscience technologies offers valuable insights into how storytelling can be strategically employed in the digital age. Further research is recommended to generalize these findings across different demographics and product categories.

P3-B-147 - Seeking or ignoring ethical certifications in consumer choice

Suzanne Oosterwijk [1], **Nicolette Sullivan** [2], **Joel Van Der Weele** [1], **Dianna Amasino** [3]

[1] University of Amsterdam, [2] The London School of Economics and Political Science, [3] Tilburg University

Details

Objective: Consumers often encounter, and claim to care about, ethical information such as the labor conditions or environmental impacts of their purchases. However, ethical information may generate a tension between instrumental information-seeking and affective ignorance, such that consumers ignore ethical information when possible because it is unpleasant to contemplate. Further, while some studies find that negative framing increases ethical considerations, affective ignorance motives suggest that it may lead to more ignorance and backfire if information is avoidable.

Methods: Across two studies, we investigated how the accessibility and framing of ethical information impacts attention and choice. Study 1 (N=152) examined incentivized choices in the lab and Study 2 (N=816) investigated hypothetical online choices in a representative sample. Participants made 42, two-alternative forced choices between consumable items (e.g., chocolate) with attributes varying in price, quality, size, and Fairtrade and Organic certifications. In Study 2, positive (or negative) framing emphasized the potential to help (or harm) workers by (not) buying Fairtrade or one's own health and environment by (not) buying Organic. Crucial to examining information-seeking and ignorance, half of the trials had "open" information with all attributes visible; the other half had avoidable "hidden" information with attributes hidden behind labeled boxes in MouselabWEB (Willemsen & Johnson, 2019).

Results: Using linear probability regressions of Fairtrade or Organic choices with clustered standard errors, we find evidence that hidden information slightly reduced certified choices, with a significant

reduction in Fairtrade choices (b=-0.05, p<0.01), but not Organic (b=-0.03, p=ns) in Study 1 and a significant decrease in Organic choices (b=-0.02, p<0.01), but not Fairtrade (b=-0.01, ns) in Study 2. Further, negative framing had a stronger impact than positive framing, regardless of open or hidden attributes (Fairtrade: negative vs. positive: b=0.05, p<0.01; Organic: negative vs. positive: b = 0.08, p < 0.001). Our MouseLabWEB attention data show that both positive and negative frames reduced ignorance compared to neutral, and negative framing increased information-seeking compared to positive framing (b=0.02, p<0.01 for both Fairtrade and Organic). Further, dwell time and ignorance moderated the use of ethical certifications in choice.

Conclusion: There is a small reduction in Fairtrade and Organic choices under hidden information, with a stronger impact of (especially negative) framing increasing ethical information-seeking and certified choices with no backfiring effect under hidden information.

P3-B-148 - Neuroforecasting demand for different types of movies

Luis Alvarez [1], **Cynthia Wu** [2], **Tara Srirangarajan** [2], **Brian Knutson** [2]

[1] Maastricht University, [2] Stanford University

Details

Neuroforecasting research suggests that a subset of the same brain regions that predict individual choice can also forecast various types of aggregate behavior. While the Affect-Integration-Motivation (AIM) framework has implicated neural targets that process anticipatory affect (i.e., the Nucleus Accumbens or NAcc in anticipation of gains and Anterior Insula or AIns in anticipation of losses) and integrative value (i.e., the Medial PreFrontal Cortex or MPFC) in predicting individual choice. According to a "partial scaling" account, however, only some of these may generalize to forecast aggregate choice out of sample (Knutson & Genevsky, 2018).

Existing neuroforecasting research has focused on markets for "goods" (e.g., internet lending rates, sharing of news article), yielding evidence that NAcc activity can forecast aggregate demand, but less research has forecast demand for "bads," or products associated with negative outcomes. Here, we used neural activity in response to movie trailers (n=30) to forecast demand for movies (n=30; indexed by first month box office returns). By incorporating two genres of trailers (i.e., comedy versus horror), we further tested whether different neural markers might forecast demand for "goods" versus "bads".

At the individual level, consistent with AIM framework predictions, peak activity in the NAcc (t=4.05, p<0.001), AIns (t=2.62, p=0.009), and MPFC (t=3.42, p<0.001) positively predicted desire to watch movies. Interactions indicated that NAcc (t=2.02, p=0.043) positively predicted desire to watch more for comedy versus horror trailers.

At the aggregate level, while sampled choice or ratings did not forecast demand, brain activity did. Specifically, while peak NAcc activity positively (t=1.87, p=0.076) and AIns activity negatively (t=−2.09, p=0.049) forecast first month box office returns, peak NAcc activity (t=2.89, p=0.009) forecast more for comedy versus horror trailers.

These findings replicate and extend previous neuroforecasting research to imply that neural markers can differently forecast demand for market "goods" versus "bads," even within the same product type, consistent with a "market matching" account. This work therefore has implications for theory with respect to inferring mechanisms that drive markets and for practice in terms of adding value to traditional forecasting methods.

P3-C-149 - A new recursive Theory of Mind model for adaptive inference about the other player's strategies

Oleg Solopchuk [1], **University Medical Center Hamburg-Eppendorf (Uke)** [1], **Giorgio Manenti** [1], **Jan Gläscher** [1]

[1] University Medical Center Hamburg-Eppendorf

Details

Theory of Mind (ToM) denotes the ability to construct a mental model of other people's latent cognitive states such as intentions, goals, and beliefs. Such models can become recursive by assuming that the other person models others too ("I think that you think that I will do X."). These different iterations are referred to as ToM levels, beginning with level 0 ("whatever you do, I will just do X"), and the other person's level is always assumed to be one level below one's own. Existing computational models have been successfully used to model choice behavior in competitive (Matching Pennies, Devaine et al., 2014) and cooperative (Stag Hunt, Yoshida et al., 2008) economic matrix games. These models typically assume a single, static reward maximizing strategy at Level-0 (e.g. Win-Stay, Lose-Shift, WSLS) and then build subsequent higher levels as best responses to the predictions of the lower level. Thus, detecting the correct strategy at the lowest level is of critical importance for building the correct ToM hierarchy. However, we know from behavioral data in a variety of economic games that WSLS is not the only strategy that participants pursue. For instance, imitating the last choice of the opponent (regardless of the outcome, i.e. Tit-for-Tat) is the most successful strategy in an iterated Prisoner's dilemma (Axelrod, 1980). Here, we develop a new computational model for recursive ToM that overcomes this limitation. Instead of assuming a single strategy, our model of Level-0 agent features a (Dirichlet) belief distribution over a finite number of strategies of her opponent, which is updated over the course of their interaction. For each strategy, the model predicts the opponent's next action, and then computes an overall belief over actions by weighting the predictions of that strategy by its belief. Subsequent higher ToM levels are constructed as best responses to the other player by marginalizing over her predicted actions. The resulting beliefs over actions are then used to update a (Dirichlet) distribution over levels, thus allowing for dynamic changes in the ToM level of the agent through the course of a matrix game. We present parameter and model recovery studies and apply the model to an exemplary dataset of matching pennies. Model comparison with existing approaches demonstrates the feasibility and importance of accounting for the variability of choice strategies in human participants.

P3-C-150 - Generalized and social anxiety exert dissociable but task-specific effects on strategic game play

Amrita Lamba [1], **Rebecca Saxe** [1]

[1] Massachusetts Institute of Technology

Details

BACKGROUND AND AIMS: Economic decision-making games offer a formalized yet rich space to identify individual differences in strategic game play, particularly those related to psychopathology. Lamba et al. (2020, 2024) showed generalized anxiety symptoms correlate with reduced learning from monetary losses in an iterated Trust Game, a finding that is selective to generalized rather than social anxiety. The interpretability of such task-parameter-symptom mappings remains unclear, however, given the limited objectives and generalizability of any individual game. The current work aims to expand on prior work by considering how model parameters relate across games with distinct learning challenges in an effort to identify stable, non-task specific individual differences in game play.
METHODS: Participants played the Trust Game (TG; N = 412) or the Asymmetric Social Exchange Game (ASE; N = 682) and were told they were engaging with other online partners (actually preprogrammed agents). In the TG, participants learned how much money to invest with three distinct partners that were initially trustworthy, untrustworthy, or neutral but subtly reversed their return rates, requiring participants to continuously adjust their investments. In the ASE game, participants placed a bid for a high or low reward (e.g., $10-$3 split) with a partner who also bids high or low. Players only earn their bids by coordinating such that one bids high while the other bids low. When partners coordinate, one partner always earns more than the other. Participants played three partner types: one who bids high, one who bids low, and one who alternates strategies. We fit one hybrid Bayesian Reinforcement Learning Utility (BRL-U) model to capture learning and social preference dynamics in both games. Model parameters were then correlated with anxiety symptoms.
RESULTS: A single BRL-U model could capture play across games. The relative learning rate, gamma, quantifies learning from negative versus positive outcomes. Gamma was positively correlated with generalized anxiety, not social anxiety, in the TG but not in the ASE game. The model parameter, alpha, adjusted the rate of learning as a function of disadvantageous inequality aversion. Alpha was negatively correlated with social, not generalized anxiety, in the ASE game but not the TG. **CONCLUSIONS:** Distinct model parameters associated with strategic choice are selectively associated with generalized and social anxiety, however, these mappings are task-specific and do not generalize across games.
ACKNOWLEDGMENTS AND FUNDING: This work was funded by the Simons Foundation and the Centers of Biomedical Research Excellence (COBRE).

P3-C-151 - The Selfishness of Adolescents Overrules Cooperation in Social Dilemmas

Xiaoyan Wu [1], **Hongyu Fu** [1], **Gökhan Aydogan** [2], **Chunliang Feng** [3], **Chao Liu** [1]

[1] Beijing Normal University, [2] University of Zürich, [3] South China Normal University

Details

Objective: Cooperation skills are crucial for individual success in society. While research has often shown that adolescents exhibit less cooperation than adults, the computational mechanisms underlying such behavioral variations remain underexplored. Existing studies commonly link adolescents' lower

cooperation to their underdeveloped mentalizing functioning, leading them to underestimate others' trustworthiness and willingness to cooperate. Here, we proposed another possibility that adolescents' lower cooperation may be due to a selfish motive to exploit others.

Methods: The present study used a repeated version of the Prisoner's Dilemma Game to compare cooperative behaviors between adolescents (N = 127) and adults (N = 134). In which, they were led to believe that they were playing a game with a human partner. Unknown to the participants, the behaviors of the partner were predetermined by a computer program. This ensured a controlled comparison of behavioral responses between adults and adolescents. We further systematically developed computational models with different assumptions to explore the computational mechanisms underlying these behaviors between age groups.

Results: Consistent with most previous studies, adolescents showed less cooperation than adults. Following the interaction of Group × Number of Previous Rounds × Partner's choice (Generalized linear mixed-effects models, $\beta_{(standardized)} = 0.24$, $p < 0.001$), we found that adolescents showed significantly less cooperation compared to adults only after the partner's cooperation but not following partner's defection. Further, adults increased cooperation in response to partners consistently cooperating, but this pattern was not observed in adolescents. We found that an asymmetric reinforcement learning within the social reward model provided the best fitting for the behaviors of both age groups. Results from the best-fitting model revealed adolescents and adults did not show a difference in building expectations of their partner's willingness to cooperate. However, adolescents exhibited a weaker social preference for cooperation, which resulted in a reduced intrinsic reward of reciprocity compared to adults. The results thus supported our hypothesis that adolescents' lower cooperation stems from a motive to exploit others rather than a deficiency in mentalizing.

Conclusions: These findings have implications for promoting prosocial behaviors and designing effective socialization interventions during adolescence by highlighting the importance of reciprocity, ultimately fostering a more cooperative society.

P3-C-152 - Persistence of strategic sophistication and transfer of learning across games

Luca Polonio [1], **Davide Marchiori** [2], **Sibilla Di Guida** [3]

[1] University of Milano Bicocca, [2] IMT school for advanced studies Lucca, [3] IMT School for Advanced Studies Lucca

Details

Study's objective

Previous research shows that under certain learning conditions, human agents not only adapt their strategic behavior but become better at thinking strategically (Marchiori, Di Guida, & Polonio, 2021). However, it remains unclear whether these behavioral changes persist over time and if an increase in strategic sophistication is confined to the original learning context or transfers to new settings. Addressing these questions is crucial, as persistent and adaptable changes in strategic behavior are essential for transferring acquired knowledge across diverse scenarios.

Methods used

To investigate this, we conducted an experimental eye-tracking study involving 240 participants engaging in a series of games with a rational, profit-maximizing algorithm. The study spanned three sessions over three months. In the first session, half of the participants (treatment condition) played different types of games with feedback to enhance their strategic sophistication, while the other half (control treatment) played the same games without receiving feedback. Sessions two (after two weeks) and three (after three months) evaluated the persistence of increased sophistication over time and its transferability from simultaneous moves 3x3 matrices to semi-sequential coordination games.

Results obtained

Using a mixed-effects logistic model on choice data, we assessed the occurrence and persistence of learning. Our results indicate that enhanced strategic sophistication persists even after three months, with a standardized odds ratio of 2.05 (95% CI: 1.68-2.50, $p < 0.001$), and the increased rate of optimal choices remains consistent across player types and initial levels of sophistication. Furthermore, we found that improvements in strategic sophistication transfer to games with different strategic dynamics, albeit to varying extents depending on player type. Additionally, analysis of eye-tracking data revealed that learning by feedback is the direct consequence of the acquisition of specific strategically relevant resolution procedures during simultaneous moves 3x3 games, which carried over to semi-sequential coordination games.

Conclusions

In summary, our findings suggest that learning through feedback is grounded in a deep understanding of the strategic context rather than simple mechanical adjustments to received feedback. This learning process entails significant modifications in how individuals acquire and incorporate information regarding players' incentives. Importantly, this learning is enduring and applicable across various contexts.

P3-D-153 - Uncovering decision strategies through fixation patterns and choice movement trajectory

Youngbin Kwak [1], **Ramiro Eduardo Rea Reyes** [1]

[1] University of Massachusetts Amherst

Details

Objective

Decisions under risk involve processing of reward and risk information. Using model classifications, we identified four distinctive decision strategies. We then corroborated their different information sampling, as well as the evidence accumulation process, by linking visual fixation to movement trajectory while making reaching actions towards choice.

Methods

Participants performed a forced-choice reaching task choosing between two non-complementary gambles. The gambles were shown as a four by four grid, with the reward and probability occupying top or bottom cells in either the right or left side of the grid, representing the left and right gamble. Participants dragged the mouse cursor to the left or right corner of the screen, while tracking eye movements and the mouse cursor position. Model classification of decision strategies was performed based on variational Bayesian inference. Analysis of fixation and mouse tracking was done by mapping each visual fixation event with the cursor position. We focused on the movement trajectory angle at each fixation event, measured relative to the direct path to the ultimate choice made. The larger the trajectory angle, the greater the tendency to move towards the unchosen gamble. This metric was used as a dependent variable in a lognormal hierarchical model assessing the impact of visual information (i.e., probability and reward).

Results

We found four dominant decision strategies: Probability Ratio, Probability Tally, EV Ratio, and SV Difference. Probability Ratio and Probability Tally only relies on probability for choice, while EV Ratio and SV Difference uses both the probability and reward information.

All strategies displayed significant shifts in trajectory angle after fixating probability from the unchosen gamble. However, only EV Ratio additionally showed a significant trajectory shift after fixating on reward.

The shift in trajectory angle in EV Ratio users was significantly different from the other three strategies, especially when the probability information from one of the two gambles strongly dominated the other. The level of shift was lower in EV ratio than the other three. In the reward domain, the level of shift was higher in EV ratio users.

Conclusions

Our results suggest differential impact of the identity and magnitude of information in evidence accumulation. Importantly, these impacts manifest divergence across decision strategies, corroborating that different strategies put different weights to the probability and reward information.

P3-D-155 - Stay, withdraw, escape: humans integrate economic opportunity costs and benefits in survival decisions

Juliana K. Sporrer [1], Ulises Daniel Serratos Hernandez [1], Jules Brochard [2], Jack Brookes [1], Samson Hall [3], Dominik Bach [3], Sajjad Zabbah [1]

[1] University College London, [2] University of Bonn, [3] Hertz Chair for Artificial Intelligence and Neuroscience, University of Bonn, Germany

Details

Objective: This study aimed to investigate how humans make embodied survival decisions in the face of economic opportunity costs.

Methods: Fifty-seven participants were instructed to collect resources (fruit on a bush) in a VR environment whilst avoiding a potential threat, an attacking apex predator (panther). The threat was designed and animated for biological realism and chased the player, while a shelter provided a place of safety. We implemented a full 6 x 4 factorial design with two independent variables: (1) threat proximity, and (2) accumulated resources. Threat proximity was implemented as distance of tall grasses from which the threat could emerge, which influences survival probability. Accumulated resources were defined by number of fruits already collected when threat appeared (i.e. the k-th fruit activated threat appearance). For risk-averse agents, greater accumulated resources translate into lower opportunity costs of withdrawal/escape. In some trials, we implemented a visual barrier (grass) in front of the bush in a way that participants could track the potential threat only from a certain position (i.e. standing position with the head up).

Results: Participants followed instructions. They collected an average of 5.5 fruits per epoch, with an 84.2% virtual survival rate. Alarm vocalizations indicated immersion within the VR environment. Prior to threat appearance, the decision to forage at all, or to withdraw to shelter immediately was determined by threat proximity. Once they started foraging, the decision to continue foraging depended on threat proximity and accumulated resources. Once the threat appeared, escape initiation time also depended on threat proximity and accumulated resources: participants were more likely to escape earlier when they had collected more fruits. Finally, participants abandoned foraging earlier when they could not continuously monitor threat appearance due to a visual barrier.

Conclusions: Human decisions to survival threats in VR appear to reflect economic decision-making in that they integrate opportunity costs and benefits. Higher probability of virtual death (threat proximity and visibility) lead to earlier withdrawal and escape. On the other hand, greater resource accumulation, and thus lower opportunity cost, also lead to earlier withdrawal and escape. This suggests that withdrawal and escape from survival threats resembles goal-directed economic decision more than an instinctive or hard-wired mechanism.

P3-D-156 - Reward Prediction Updates: Tracking Changes in Reward Expectations

Stefan Bucher [1], **Zih-Yun Yan** [2], **Bo Shen** [3], **Peter Dayan** [4], **Paul Glimcher** [2]

[1] Massachusetts Institute of Technology, [2] New York University, [3] NYU Grossman School of Medicine, [4] Max Planck Institute for Biological Cybernetics & University of Tübingen

Details

OBJECTIVE: It is widely established that reward prediction error signals can be observed in the human brain: When participants believe there is a given percentage chance of earning a reward and then receive it, areas such as the ventral striatum and the ventromedial prefrontal cortex show BOLD activations reflecting the reward prediction error, relative to the reward expectation. This study aims to determine where and how these reward expectations are updated. Specifically, it asks whether BOLD signals in the ventral striatum and ventromedial PFC, known to encode reward prediction errors (Schultz, Dayan, and Montague, 1997; Caplin et al., 2010; Rutledge et al., 2010; Niv et al., 2012), also reflect updates to a participant's reward expectations *in the absence* of actual rewards.

METHODS: In an fMRI experiment, participants (N=14) faced binary lotteries represented as partially occluded pie charts. The pie chart's composition on each trial was revealed only gradually to participants, so that they could sequentially update their belief about the probability of winning the prize. At a random time during this gradual resolution (and hence at different levels of ambiguity), we elicited participants' valuation of the current pie chart-lottery using an incentive-compatible procedure yielding their probability equivalent.

RESULTS: In line with theoretical predictions, participants' valuations in absence of ambiguity reveal (1) that their subjective beliefs closely track the lottery's winning probability, and (2) their degree of risk aversion, with a coefficient of relative risk aversion of 0.93. In the presence of ambiguity, participants' revealed beliefs are slightly more pessimistic, reflecting their degree of ambiguity aversion. Relating the neuroimaging data to behaviorally revealed belief updates, we then examine the extent to which BOLD activations in reward-prediction areas track changing reward expectations.

CONCLUSION: This study examines whether areas encoding reward prediction errors in response to received rewards also represent updates to reward expectations, mirroring punishment expectation errors seen in similar locations (Seymour et al., 2003). Promising a measurable trace of belief updates, this approach offers novel empirical constraints on models of hitherto unobservable evolving beliefs, and provides a rigorous way to test neuroeconomic theories that make joint predictions about choice behavior and neural data.

(S.B. & Z-Y.Y. co-first authors, P.D. & P.G. co-last authors)

P3-D-157 - Bayesian-Optimality of Decision-Making Behavior Across Time

Mathi Manavalan [1], **Iris Vilares** [2]

[1] University of Minnesota, [2] University of Minnesota, Twin Cities

Details

Objective: Bayesian Decision Theory (BDT) has been used extensively to model human decision-making behavior, particularly within sensorimotor contexts, and has shown that behavior is quite similar to what it considers to be optimal. BDT describes optimal decision-making behavior as combining prior with current sensory information according to their relative uncertainties. One paradigm in particular, called the coin-catching task, has been shown to produce qualitatively Bayesian-like behavior and has recently been used to investigate behavioral differences between those with psychiatric conditions and

neurotypical populations. But less is known about the test-retest reliability of the task, which is necessary to be able to characterize individual differences and assess trait-like behavior. Here, we investigate this through a test-retest study of the coin-catching task. We also assess the extent to which behavior can be considered as Bayesian-optimal.

Methods: We used a coin-catching task, a paradigm in which two levels of uncertainty are imposed on prior and current sensory information. Participants (N=103), recruited through the university system, completed the task at two time points which were at least two weeks apart. Participants conducted the coin-catching task in which they had to guess the location of a hidden coin on each trial. They could not see the coin but could see the splashes it made. The task had 480 trials split into 4 blocks. We varied the prior information uncertainty by changing the variance of the distribution from which the coin location is drawn, resulting in either low or high prior uncertainty (alternated across blocks). The variance of the current sensory information was manipulated by changing the spread of the splashes, resulting in low or high sensory uncertainty (changed trial-by-trial).

Results: We found that participants made decisions in a qualitatively Bayesian-like manner, relying more on information that is less uncertain. We also found high test-retest reliability of the task, ($ICC[3,1] = 0.688$, $p < .001$). The average relative weight on sensory information was found to be highly variable across participants (with average relative weights ranging from -.152 to .935), but highly correlated across time points within participants ($r(101) = 0.69$, $p < .001$).

Conclusions: Our results indicate that the coin-catching sensorimotor task can be reliably used to study individual differences in decision-making behavior. Furthermore, our results suggest that the way someone combines prior and current information may be a character trait given the strong correlation within individuals.

P3-D-158 - Evaluating the Test-Retest Contextual Reliability of Novel Metrics of Metacognition and Value-Based Choice

Morkeh Blay-Tofey [1], **David Quezada** [1], **Parmis Khosravi** [1], **Corey Ziemba** [2], **Daniel Pine** [1], **Silvia Lopez-Guzman** [3]

[1] National Institute of Mental Health, [2] University of Texas at Austin, [3] National Institutes of Health

Details

Objective: Metacognition is a higher-order cognitive ability that is impaired in many mental health disorders. It is addressed in computational psychiatry through models that measure confidence in tasks requiring perceptual judgments. However, the existing models are not suitable for value-based decisions, where choices are neither correct nor incorrect, but represent an implicit preference. We introduce a novel task and computational framework, estimating individual preferences for uncertainty (e.g., risk or ambiguity aversion) and confidence bias and sensitivity. Our task demonstrated test-retest reliability across different sessions and settings in healthy adults (ages 18-55). We then used raw behavior to model confidence via the Confidence as a Noisy Decision Reliability Estimate (CASANDRE)

Model. This model applies Signal Detection Theory, where the subjective value difference (SVD), known as the signal, is normalized by noise. The precision of the noisiness estimate is a parameter in the confidence model.

Methods: N = 22 healthy adults participated in two sessions, one in a lab and one at home, with settings counterbalanced. Decisions involved choosing between certain monetary outcomes and probabilistic lotteries, with varying amounts, probabilities, and ambiguity levels across both gain and loss trials. Participants rated their confidence on a Likert scale. Computational modeling yielded parameters for risk tolerance, ambiguity tolerance, and choice stochasticity in a utility model where reliability was assessed by comparing the parameter values across environmental and temporal contexts. This comparison involved finding the average parameter value across subjects and comparing across contexts. Confidence reliability was assessed by median splitting on SVD to separate “easy” or high SVD trials from “low” trials, finding the average confidence, and comparing across contexts.

Results: Intraclass correlation coefficients (ICC) for utility parameters indicated good reliability (0.6-0.7), with confidence ratings ranging from 0.5 to 0.8 across different contexts.

Conclusions: Preliminary findings suggest our measures of choice and confidence show good stability even across different contexts. We aim to obtain confidence reliability from the CASANDRE model in adults and extend this study to children and adolescents, focusing on risk and confidence to evaluate the task's reliability in that age group. Data collection is ongoing for adults and is starting for children and adolescents. We plan to discern whether uncertainty and confidence modeling is reliable within adults and contrast this against the data and analysis of children and adolescents.

P3-D-159 - Evaluation of cognitive aspects in decision-making under risk: an eye-tracking study

Anderson Lucas Carneiro De Lima Da Silva [1], **Fabiana Bispo Da Silva** [1], **Flávio Lopes Da Silva Filho** [1], **Jefferson Souza Medeiros** [1], **Maria Clara De Oliveira Gê** [1], **Ana Paula Cabral Seixas Costa** [1], **Lucia Reis Peixoto Roselli** [1], **Danielle Morais** [1], **Adiel de Almeida** [1]

[1] Universidade Federal de Pernambuco

Details

Decisions often have to be made in scenarios permeated by risk. Two main factors that influence the way choices are made in risk contexts are probability and preference for the desired payoff.

Objective: This study follows in the line of risk assessment and decision-maker (DM) behavior (Lai & Huang; 2019; Dertwinkel-kalt et al, 2023), exploring cognitive aspects related to attention and cognitive effort and how they are employed in different conditions: gain and loss.

Method: An experiment was conducted using the Eye-Tracking Tobii X-120 through which fixation and pupil size data were collected. The stimuli consisted of lotteries compared with safe options in gain and loss situations. In both situations, the probability was 50% in each payoff. In the gain stage, the payoffs

were 1000 BRL and 0 BRL, while in the loss stage, they were 0 BRL and -1000 BRL. At each iteration, the value of the certainty equivalent varied until an indifference was identified. The experiment was carried out with 30 subjects at the Laboratory for Neuroscience and Behavioral Studies in Decision (NSID). Regression and correlation analyses were carried out, as well as the paired T-test.

Results: The analyses showed that fixations were higher when viewing the lottery compared to the safe option. Fixations were also higher in the loss condition than in the gain condition for the lottery. These results suggest that participants are more sensitive to probability and return under risk, especially when faced with losses. On the other hand, pupil size was greater in the gain condition than in the loss condition, which may suggest a lower engagement behavior in loss conditions. The results indicated that attention to the lottery increases with age, and the more risk-averse they were, the more they fixated on the probabilities and payoffs in the gain situation. In addition, a positive correlation was found between the certainty equivalent and cognitive effort, as well as between the certainty equivalent and the level of attention. The gender of the subjects was not a significant variable in this study.

Conclusion: In this sense, the results suggest a relationship between cognitive effort and attention with risk assessment along the lines of Prospect Theory (Kahneman and Tversky, 1979), as well as the impact of age on behavior, complementing the existing literature in the area. Finally, it is hoped that this research will contribute to establishing strategies in risk situations when there are gains and losses involved, based on consideration of the underlying cognitive aspects.

PS.03.01, P3-D-160 - The anterior insula encodes salience in a subjective fashion and preferentially in the aversive domain

Jae-Chang Kim [1], **Stephan Nebe** [1], **Lydia Hellrung** [1], **Philippe Tobler** [1]

[1] University of Zürich

Details

The insula has been associated with processing salience (Molnar-Szakacs & Uddin, 2022) and salience prediction errors (Corlett et al., 2022). However, the exact form (e.g., objective vs. subjective) and extent (e.g., preferentially in aversive domain or domain-general) of salience processing by the insula remains to be determined.

To address these questions, we used a Pavlovian conditioning task (n=41; 22.4 ± 0.43 years, mean ± SEM; 19 women) involving appetitive, aversive, and neutral liquids. We behaviorally matched the absolute value of appetitive and aversive liquids by bidding (Becker-DeGroot-Marschak auction) and rating (general labelled magnitude scale), and used different cues to predict the liquids with different probabilities (p=0, 0.5, 1). In each trial, participants rated both the cue and the outcome, which allowed us to directly determine subjective salience prediction errors, the unsigned difference between the outcome rating and the cue rating within the MRI scanner. By contrast, objective salience prediction errors were defined by probabilities predicted by cues and actual outcomes.

A non-parametric analysis revealed a significant objective salience prediction error signal for valenced outcomes (i.e., appetitive and aversive) but not for non-valenced liquids (i.e., neutral) within the insula

at the time of the outcome ($p<0.05$, FWE-whole brain voxel-level corrected). A parametric approach replicated these results but direct comparisons revealed a significantly stronger association of anterior insula activity with subjective than objective salience prediction errors at the time of the outcome. Moreover, subjective salience prediction errors activated the anterior insula also at the time of cues. As one would expect based on discounting, the cue-related signals were significantly weaker than the outcome-induced signals. Finally, separate analyses for the different domains (i.e., appetitive, aversive, neutral, and no outcome) revealed domain-general subjective salience prediction error signals in the anterior insula at the time of the outcome. By contrast, at the time of the cue, the coding of subjective salience prediction errors was limited to the aversive domain, compatible with weaker discounting of salience in the aversive than the appetitive or neutral domain.

Together, the anterior insula encodes an error in the prediction of subjective salience at each moment in time and discounts future salience to a lesser extent in the aversive than the appetitive domain. These findings reconcile two hitherto apparently conflicting views on insula function (salience coding common to appetitive and aversive domain vs. preferential coding of aversive events).

P3-D-161 - Biased processing of multiple outcomes in human reinforcement learning: evidence from computational modelling and eye-tracking

Henri Vandendriessche [1], **Charlotte Gruson** [2], **Antonios Nasioulas** [3], **Camille Straboni** [4], **Stefano Palminteri** [1], **Maël Lebreton** [3]

[1] École Normale Supérieure, [2] Université Paris Nanterre, [3] Paris School of Economics, [4] Ecole Normale Supérieure

Details

Objectives

Standard reinforcement learning tasks focus on situations where each choice delivers only a single outcome at a time. However, in many circumstances choices result in multiple simultaneous outcomes that have to be integrated in order to learn correct estimations of reward expectation.

Methods

In this study, we adapted the classical two arm bandit task used in reinforcement learning experiments by manipulating the number of outcomes delivered by each option: in certain learning contexts, the choice of an option leads to two outcomes instead of one. The two outcomes were drawn from the same outcome-specific underlying distribution (which represented the true expected value of the option), meaning that both outcomes, even if they can be different on a given trial, were equally informative for the sake of learning. However, the two outcomes were not equally relevant in terms of reward, as a random selection process selected one of the two outcomes to count for the final payoff of the participant. We added a second factor in our now, 2 by 2 orthogonal design, which is a reversal of contingencies at mid session. The goal here is to assess the impact of two outcomes over one, in the post reversal recovery of option values.

Results

Behavioural results (*n=36*) show that first: both factors, number of outcomes ($\chi^2(1,36)=6.81, p=0.009$) and presence of a reversal ($\chi^2(1,36)=41.25, p<.001$) are statistically significant. Second, probability to stay or switch from option in two-outcomes conditions, is highly influenced by outcomes from the last trial ($\chi^2(3,36)=109.96, p<.001$). Indeed, both outcomes are considered, however they are not treated symmetrically as function of their valence. The pattern is even more striking in incoherent situation where both outcomes are different (one positive and one negative). Participants tend to integrate information in a positive way putting more value on positive than negative outcomes. This behavioural observation is backed-up by computational analyses showing that, on top of previously documented asymmetric update, subjective outcome encoding is biased, such as it overweights rewards over punishments. Behavioural and computational results were paralleled by eye-tracking analysis showing that attention deployment was biased by outcome valence and relevance especially in incoherent situations, where two outcomes would differ in valence. The main results were confirmed in a second experiment (*n=39*) featuring full feedback information.

Conclusion

Overall, when options deliver multiple outcomes, we found behavioural, modelling and visual attention traces, showing that they're integrated in an optimistic and confirmatory manner.

P3-E-162 - Top-down and bottom-up motivations for self-control exert unique influences on nutritional decision making

Matthew Bachman [1], **Sophie Faessen** [1], **Cendri Hutcherson** [1]

[1] University of Toronto

Details

Introduction: Rising obesity rates have spurred the search for strategies that can encourage healthier food choices. Several papers have shown that top-down cognitive regulation strategies can help people incorporate healthiness more into their choices (e.g., Hare et al., 2011). However, the efficacy of these strategies is undermined by recent work suggesting that these effects are fleeting (Wilson et al., 2023), and may not modulate underlying neural value signals (Tusche & Hutcherson, 2018). Yet there other processes, such as the physiological state of satiation, that can alter the valuation of foods in favor of healthy eating. However, it is unclear whether such bottom-up changes in food preference operate through similar or different processes. Methods: We designed a study capable of testing the separate and combined influences of extrinsically motivated, top-down regulation and intrinsically motivated, bottom-up self-control during food choice. Hungry participants (current n = 29 out of 50) came to two separate study sessions. They first rated the tastiness and healthiness of different foods before completing a dietary choice task. On each trial they used a mouse to indicate their preference between a displayed food or neutral, default food. Top-down motivations were manipulated in different blocks by instructing participants to either choose naturally or to focus on the food's healthiness. Bottom-up motivations were manipulated between sessions by testing participant's food choices while either hungry or just after gorging on a buffet of junk food. Results & Discussion: Our results currently show

that hungry participants naturally used taste but not health to make their choices. Top-down regulation reduced the weight given to taste and increased the incorporation of health. In contrast, sated participants naturally down-weighted taste and incorporated health. The influence of health on choice was even larger when sated participants regulated their choices, suggesting that these different motivations provide different but additive benefits. Analyzing the influence of taste and health on mouse trajectories yielded additional insights into the dynamics of these different motivations. Satiation resulted in a smaller but faster inhibitory influence on taste, while top-down regulation generated a slower but stronger influence. Our results suggest that top-down and bottom-up motivated self-control may exert distinct influences on choice processes. We have initiated a parallel fMRI study, and aim to have preliminary results on how these different motivations alter representations of taste and health in the ventromedial and dorsolateral prefrontal cortex.

P3-E-163 - Age differences in integration of values for self and others: Protective role of inhibitory control

Anita Tusche [1]

[1] Queen's University

Details

While aging societies worldwide represent significant social and economic challenges, there are also opportunities. For instance, past research suggests higher prosociality in older people compared to younger adults, a vital component for the functioning of societies. However, recent meta-analytical evidence indicates that these effects are heterogeneous, may be small, and are influenced by how prosociality is measured. Further, the precise cognitive and computational factors contributing to age-related differences in prosocial behavior remain largely unknown. Combining a value-based decision framework with computational modeling, our study sheds light on differences in how older and younger individuals solve the cost-benefit tradeoff of (pro)social decision problems. Using Bayesian hierarchical drift-diffusion modeling and data from a modified dictator game, we investigated prosocial decision-making in younger (n = 63) and older adults (n = 48). We revealed differences in how older and younger individuals incorporate information about potential gains and costs for themselves and others: Younger adults integrated values for benefits for themselves and others in the decision-making process, explaining when and why people act prosocially. However, OA showed improved decision-making efficiency when solely considering values for self and others separately. Interestingly, individual differences in inhibitory control in older adults moderated the age effects: older adults with stronger inhibitory control abilities made decisions based on the integrated information of benefits for themselves and others, showing a behavioral pattern like younger adults. Our findings emphasize social decision-making as a cost-benefit analysis that is solved differently by younger and older adults, providing novel insights into the mechanisms of how the social decision process changes across the lifespan. The results also point to the precise way via which inhibitory control moderates age-related differences in social decision-making, shedding light on the heterogeneity of individual aging trajectories. The discussion will bear on the broader idea of possible alternative pathways, e.g., by considering metastability of the aging brain and state changes in dynamical systems that constantly occur to successfully cope with the molecular, epigenetic, neural, social, or contextual changes that

occur over a person's life course. These findings offer insights into the behavioral and computational mechanisms influencing healthy aging.

P3-E-165 - Choice Strategies in Sequential Decision-Making

Sergej Golowin [1], **Christoph Korn** [2]

[1] Heidelberg University Hospital, [2] University of Zürich

Details

Study objective

The quality of our current decisions determines our future choice options. This temporal dependence of decisions fulfills the so-called Markov property. We developed a sequential decision-making paradigm that fulfills this Markov property. Our task is designed to detect decision strategies based on ecologically valid factors. These factors are chosen to mimic behavioral contexts related to the approach and avoidance conflict. We combine behavioral data with computational models (which use approaches from reinforcement learning and supervised learning) to test and compare choice strategies under different environmental/experimental conditions.

Methods

We developed a hunter-gatherer game, in which healthy participants (n = 29) had to make binary choices with probabilistic outcomes. The choices were designed to emphasize two opposing strategies in two experimental conditions: maximizing potential rewards (approach environment), or minimizing predation risk (avoid environment). Our task allowed to test different decision models of varying complexity, including a number of single- and multi-heuristic policies, and the optimal strategy (computed from a fully observable Markov decision process). Due to the binary response variable of the task, model comparisons could be provided via logistic regression and Bayesian model inference.

Results

Participants generally use a mixture of multiple heuristic policies (protected exceedance probability *PEP* = 0.98; range β_1 = [-0.08, 10.57], range β_0 = [-3.65, 1.41]). However, if the environment emphasizes more avoidant strategies, choices become more optimal (optimal strategy: *PEP* = 0.47; range β_1 = [-0.60, 30.54], range β_0 = [-1.66, 1.44]; multi-heuristic strategy: *PEP* = 0.45; range β_1 = [-0.54, 11.81], range β_0 = [-3.98, 1.63]). Based on this, it seems like some individuals are triggered to behave more optimal by changing their strategy from approaching an environment to avoiding it.

Conclusion

We interpret our results in terms of an internal trade-off between efficiency and accuracy of behavior. We argue that the balance between more efficient and more accurate decisions is a key drive in behavioral ecology. We draw a comparison to comparable trade-offs like model-free vs. model-based

decision-making, and automatic vs. deliberate behavior. By these means, we approach an ecologically valid understanding of dynamic and sequential choice. This bares the potential to render an evolutionary context for seemingly paradoxical choices in behavioral economics. Furthermore, our task design allows studying the neural principles of ecologically immersed cognition and their implications for mental health in the future.

P3-E-166 - Risk attitudes in monetary lottery task and in an equivalent time lottery task

Wen-Yi Yang [1], **Shih-Wei Wu** [1]

[1] National Yang Ming Chiao Tung University

Details

Many decisions involve evaluating options with varying degrees of risk or uncertainty. In some scenarios, such as financial investments, risks are primarily associated with monetary outcomes—the probability of losing or gaining money. In other instances, risks are tied to the timing of rewards, such as the duration one might wait to be seated at a restaurant. Although prior research has compared risk preferences with time preferences, few studies directly compared risk attitudes toward monetary outcomes with those toward time delays. In this study, we developed a novel experimental paradigm where subjects engaged in a standard monetary lottery task (choosing between lotteries with monetary gains) and a mathematically equivalent “time” lottery task (choosing between lotteries with time delays as outcomes). The experiment consisted of three sessions. Subjects first performed an intertemporal choice task that we used to estimate temporal discount function. We then used the subject-specific temporal discount function to create time lotteries that were mathematically equivalent to monetary lotteries. Subjects then performed monetary lottery and time lottery tasks in separate sessions, with order randomized across participants. We adopted a well-established “common-consequence” lottery design to quantitatively characterize risk attitudes and non-parametrically examine probability weighting. At the group level, we found that participants (n=30) on average were more risk averse in the time lottery task than in the monetary lottery task. For probability weighting, subjects exhibited moderate probability distortion in both tasks. At the individual level, only 13 out of 30 subjects showed significant correlation in choice behavior between the two tasks. Together, these results suggest that risk attitudes depend on the domain of outcomes in risky prospects (money or time) and that individuals do not have consistent risk preferences across these different domains.

P3-E-167 - Improving self-control: the effect of role models’ observation on near and far transfer

Mina Movahedi [1], **Gayannee Kedia** [1], **Katja Corcoran** [1], **Amelie Baitinger** [1], **Nussgraber Clemens** [1]

[1] University of Graz

Details

Self-control leads to a happy, healthy life. Scholars have proposed different methods to improve self-control, but all these methods have one major drawback: they are effortful and thus ill-suited for people with low self-control. The present studies suggest a less effortful method: Observational learning.

Previous research has shown that observing a role model exerting self-control can improve observers' self-control on the same task (near transfer). But is it effective on a different task (far transfer)? We try to answer this question in three studies.

In these studies, we exposed participants to the intertemporal choices (e.g., choosing between $25 now and $55 in a week) of a role model and assessed their self-control in the near and far transfer tasks.

In Study 1 (N=270), we found a significant influence of the role model on the near transfer task and a tendency for the far transfer task, but this latter result was marginally significant. In Study 2 (N=279), we made the role models more inspiring and found a significant influence for both near and far transfers. In Study 3 (N=179), we repeated the previous studies in the lab recruiting low self-control participants. We also used a new far transfer task. We found a significant influence of role model observation on the near transfer task, but we didn't find the effect on the far transfer task.

The results demonstrate that observing a role model's self-control positively impacted participants' self-control in the near transfer tasks. However, despite efforts to enhance role model inspiration, the effect on the far transfer tasks remained inconclusive.

P3-E-168 - Developing an ontology for quantifying risk of transitioning from occasional opioid use to opioid use disorder

Aysu Secmen [1], **Sage Harley** [2], **Paul Glimcher** [1]

[1] NYU Grossman School of Medicine, [2] New York University

Details

AIM: Despite widespread opioid use at a societal level, only a minority of individuals exposed to opiates progress to opioid use disorder (OUD). Can individuals predisposed to developing OUD upon opiate exposure be identified before first use? In this initial study, our objective was to develop an instrument that could accurately differentiate cohorts with different opiate-use statuses, laying the groundwork for predictively differentiating those at risk of OUD in a comprehensive, large-scale longitudinal study.

METHODS: We collected data for 27 well-validated instruments (550 individual questions) covering four risk domains: Overall Life Quality, Opioid-Induced Hedonic Experience, Genetic Predisposition, and Psychological Predisposition. We had 150 participants equally divided between three matched cohorts: those with a history of OUD, those exposed to opiates but with no history of OUD, and a control cohort with no exposure. Employing a two-step modeling process, we applied dimension reduction methods to our dataset comprising 550 data points per participant, and utilized linear discriminant analysis for

classification. Our goal was to identify the minimal number of questions required to classify these cohorts.

RESULT: We were able to identify just 10-25 questions extracted from the 27 instruments that were, when combined together, able to separate the three cohorts with reasonable efficiency. Questions related to lifetime stress, coping, social support, reward responsiveness, and impulsivity emerged as influential in separating the cohorts. Accuracy rates were in the 80%-94% range for the training, and 70%-85% range for the testing data, for these 10-25 questions.

CONCLUSION: Using a between-subjects approach as an initial step, we were able to identify opiate-use status with high precision. This lays the groundwork for employing these features in a longitudinal predictive study. Such an approach holds promise for accurately identifying individuals at risk for OUD prior to first exposure, thereby contributing to the advancement of early intervention strategies.

P3-F-169 - Acting in social groups can induce an inflated sense of confidence

Yongling Lin [1], **Tianqi Zhan** [1], **Stephen K Fleming** [1], **Marco K Wittmann** [1]

[1] University College London

Details

Social context exerts powerful influences on individuals' confidence. For instance, the 'Bay of Pigs Invasion' in 1961, one of the biggest failures in US military history, is thought to be caused by poor group decision-making paired with an inflated sense of confidence. Previous studies have examined how social influence and advice-taking affect individuals' confidence. However, little is known about how the mere knowledge of others' performance can affect how we judge ourselves. To address this question we conducted two studies with a combined total of 181 participants, utilizing a new, precisely controlled 4-player social working memory paradigm. On every trial, participants observed quantitative performance sequences about themselves (S), a partner (P), and two opponents (O1, O2). Afterwards, participants compared the scores of selected players from the two teams from memory (e.g., S vs O1, P vs O2) and decided on the better performer, before providing a confidence rating in their decisions. Strikingly, study 1 (N = 55) revealed that participants did not fully ignore irrelevant players: When both the participant (S) and their partner (P) performed well, participants were more likely to rate their relevant own-team member as performing better than the relevant opponent, despite only one own-team member being "decision relevant" ($t(54) = 3.896$, $p < 0.001$, Cohen's d = 0.525). We found that such 'self-other mergence' translated into an inflated sense of confidence: participants were more confident if both themselves and their partner performed well ($F(1,54) = 21.836$, $p < 0.001$, $\eta^2 = 0.171$). For example, participants felt more confident in judging that they scored higher than an opponent (S>O), when an irrelevant partner (P) also scored higher, despite the fact that the rational decision variable in this task did not comprise P. These findings were replicated in study 2 (N = 126): a significant self-other mergence effect was observed ($t(125) = 9.003$, $p < 0.001$, Cohen's d = 0.802), contributing to heightened

confidence ($F(1,125) = 81.572$, $p < 0.001$, $\eta^2 = 0.247$). These results demonstrate that an inflated sense of confidence in our decisions can emerge from mere knowledge of others' performance, even if they are not relevant to the task at hand. More broadly, our results indicate that confidence may be driven by a readout of the consistency of habitual choice patterns– including their subtle social distortions – rather than reflecting a normative probability of being correct.

P3-F-170 - Communication with Surprise – Computational and Neural Mechanisms for Non-Verbal Human Interactions

Tatia Buidze [1], **Tobias Sommer** [2], **Ke Zhao** [3], **Xiaolan Fu** [3], **Jan Gläscher** [4]

[1] University of Hamburg, [2] Hamburg University, [3] Chinese Academy of Sciences, [4] University Medical Center Hamburg-Eppendorf

Details

Communication, typically based on mutual expectations, encounters difficulties in the absence of a shared language. Our study investigates how individuals naturally leverage fundamental principles of the physical world to overcome communication barriers. This scenario can occur when members of different tribes, familiar with basic non-verbal cues but lacking a common language, attempt to communicate. Here we propose that in such scenarios, intentional use of unexpected events, or surprise, can effectively communicate goal-relevant information by defying expectations that are rooted in universal shared knowledge. We test this hypothesis in the context of Tacit Communication Game (TCG), a non-verbal communication game played on a square grid board, where the Sender must convey the Receiver's goal location through her movement patterns ('messages'). By observing the message, the Receiver infers his goal location. We developed a novel computational model for the Sender's message design that is based on the idea of communicating through surprise: it uses intuitive priors based on principles of movement kinetics and goal orientation and constructs messages step by step by maximizing information-theoretic surprise at the Receiver's goal state. We evaluated our model against two simplified versions, each incorporating only one of the key components: principles of movement kinetics or goal orientation. We then fitted these models to behavioral data collected from two datasets, involving 29 and 31 pairs of participants respectively, who played the roles of Sender and Receiver in the Tacit Communication Game (TCG). The participant's behavioral data (message type, message profile, i.e., the number of moves in either direction) was better described by the full surprise model than the one component models, indicating that both universal principles of movement and goal information are important to communicating the intention successfully. The effectiveness of the surprise model was further confirmed, as its central variable—information-theoretic surprise—accounted for the Receiver's physiological and neural responses. Specifically, we found significant correlation of model derived surprise variable with the pupil diameter, indicating cognitive effort, and neural activity in brain areas related to expectancy violations, such as anterior cingulate cortex (ACC). Our results indicate that in a novel environment, surprising events can be intentionally deployed to guide the Receiver's attention and goal orientation, and therefore convey the communicative intention of the message more effectively.

P3-F-171 - Part of my pack? Examining How Affiliation Shapes Competition and Social Hierarchy Learning

Rémi Janet [1], **Caroline Charpentier** [2], **Andy Stokes-Noonan** [1], **Esther Diekhof** [3], **Anita Tusche** [1]

[1] Queen's University, [2] University of Maryland, College Park, [3] Universität Hamburg

Details

Background: Social hierarchies are inherent across species, rapidly emerge, and shape power dynamics within groups. Understanding one's position within a social setting is crucial to avoid harmful social interactions. People can learn social hierarchies during competitive interactions, showcasing individual differences in learning aptitudes and competitive behaviors. However, humans navigate multiple social groups. Yet, little is known about how affiliations with different groups and identities affect competitiveness and social hierarchy learning. **Objectives:** The current study tested how social group affiliations and personality traits influence social learning and competitive behaviors. **Methods:** Participants (Canadian hockey fans, n=67) engaged in a competitive social learning task with opponents whose skills (superior, intermediate, inferior) they needed to learn. Participants interacted with members of two groups: an Ingroup (fans of their favorite hockey team) and an Outgroup (supporters of their least favored team). To capture individual differences in competitive behavior and social learning, we used reinforcement learning models to estimate variations in choice parameters as a function of group affiliation (In/Outgroup) and personality traits (social dominance orientation, SDO). **Results:** Individuals exhibited heightened competitiveness when facing fans of their own team (Ingroup). The effect of group affiliation on competitiveness was modulated by participants' dominance orientation attitudes, with stronger effects among individuals with stronger preferences for hierarchical social structures. Parameters of the best-fitting computational model also revealed differences in learning from victories and defeats in social hierarchical learning, irrespective of whether they interacted with opponents from the In- or Outgroup. **Discussion:** The findings reveal enhanced ingroup competitiveness, potentially indicating efforts to establish social rank within one's Ingroup. This reflects the importance of social identity in shaping competitive behaviors and might suggest that losing against an outgroup member is more aversive than losing against an ingroup member. Furthermore, predispositions toward anti-egalitarianism, as measured by SDO scores, modulate competitive behaviors, emphasizing the influence of personality traits on individual actions within a social context.

P3-F-172 - Ventromedial prefrontal cortex lesions disrupt learning to reward others

Margot Gueguen [1], **Jo Cutler** [1], **Daniel Drew** [2], **Matthew Apps** [1], **Deva Sanjeeva Jeyaretna** [2], **Masud Husain** [2], **Sanjay Manohar** [2], **Patricia Lockwood** [1]

[1] University of Birmingham, [2] University of Oxford

Details

Background. Reinforcement learning is a fundamental process for how humans and other animals attain rewards for themselves. However, to act prosocially, we must also learn how our choices reward others.

The ventromedial prefrontal cortex (vmPFC) has been independently linked to reinforcement learning and prosocial behaviour, yet its causal impact and the roles of its diffuse subregions remain unknown.

Objective. The current study examined how focal vmPFC damage affects learning to benefit oneself, another person (prosocial learning), or neither person (control), and the role of mPFC subregions. We adopted a computational neurology approach combining computational modelling and lesion symptom mapping.

Methods. A large cohort of adults with rare focal vmPFC damage (n=28), and two carefully age- and gender-matched control groups (lesions elsewhere, n=21; healthy controls, n=124) completed a probabilistic reinforcement learning task where they learnt to obtain rewards for another person (prosocial), themselves (self), or neither individual (no one, control condition) on separate trials.

Results. A novel computational model incorporating separate learning rates for positive and negative prediction errors (PEs) best explained behaviour in all groups. Importantly, patients with vmPFC damage were less accurate and had lower learning rates from positive PEs when rewarding another person relative to when no one benefitted (accuracy OR=0.78, SE=0.07, Z-ratio=-2.89, FDR-adjusted p=0.017; learning rates from positive PEs OR=-0.11, SE=0.04, t-ratio=-3.06, FDR-adjusted p=0.004). They also exhibited higher learning rates for negative PEs when learning for others relative to self (OR=0.06, SE=0.03, t-ratio=2.85, FDR-adjusted p=0.023), compared to controls. vmPFC patients additionally showed a reduced self-benefitting advantage. They were equally accurate and learnt at a similar rate from positive PEs for self and neither individual (accuracy OR=0.96, SE=0.08, Z-ratio=-0.50, FDR-adjusted p=0.65, BF01=3.61; learning rates from positive PEs OR=0.02, SE=0.04, t-ratio=0.53, FDR-adjusted p=0.60, BF01=3.39). Strikingly, voxel-based lesion-symptom mapping revealed that damage to subgenual anterior cingulate cortex (sgACC) and anterior cingulate cortex gyrus (ACCg) specifically disrupted prosocial reinforcement learning.

Conclusion. These findings highlight the importance of vmPFC integrity for multiple aspects of reinforcement learning, with damage to sgACC and ACCg critical for learning to reward others.

P3-F-173 - The Influence of Peer Victimization on Social Decision Making

Sarah Salzgeber [1], **Ana Cubillo** [2], **Todd Hare** [3]

[1] University of Zurich, [2] Jacobs Foundation, [3] University of Zürich

Details

Peer victimisation experiences, during childhood, are a risk factor for mental health problems. While a broad body of research has investigated the impact of childhood stress on structural and functional brain connectivity measures, little is known about the effects of childhood stressors on decision behavior and its underlying neural mechanisms. This project aims to investigate how late childhood/adolescent peer victimization affects social decision making and associated neural activity.

Here we present results from a pilot sample of 48 healthy young adults. We will extend these analyses to test for group differences in a set of young adults (21 years) drawn from an ongoing longitudinal

sample with different levels of peer victimization experiences between the ages of 11 and 17 years (N=138). We used a modified Punishment game (with Dictator game control trials) to engage emotion and (social) decision-making, systems, which have been shown to be sensitive to early life stressors. In the task, participants choose between two payoff options, each providing a reward for themselves and for another player (Player 2). After making a choice, subjects received monetary punishments and/ or emotional feedback from Player 2.

A Bayesian mixed effects regression on choices revealed that subjects were more likely to make prosocial choices in the Punishment condition compared to the Dictator condition (post. prob > .99), and even more so if they had learned that the current partner tended to punish strongly (post.prob>.99). Beyond monetary punishments, subjects were sensitive to the partner's emotional state. In the Punishment condition, prosocial behavior increased for angry and sad emotions (post.prob>.99), while in the Dictator condition prosocial behavior decreased for angry (post.prob=.97). Using a univariate fMRI analysis, we find strong differences in the dorsal anterior cingulate cortex between the Punishment and Dictator conditions at the time of feedback on punished compared to unpunished trials ($p<0.05$ FWE). Furthermore, a representational similarity-based multivariate analysis allowed us to predict subjects' prosocial tendencies above chance (57.3%). Feature importance scores from a Random Forest classifier highlighted the limbic system, salience network, ventral attention network, and control networks in the cortex, as well as the putamen, parahippocampal gyrus, subthalamic nucleus, and globus pallidus in the subcortex, as critical regions.

In summary, our pilot data reveal robust links between behavioral and neural patterns within a social decision-making paradigm, which will be further tested in the larger sample to assess the impacts of childhood peer victimization.

P3-F-174 - Human Detection and Alignment with Large Language Models' judgements in Moral Decision-Making Context

Stefano_ Palminteri_ [1], **Basile Garcia** [2]

[1] École Normale Supérieure, [2] University of Geneva

Details

Background:

Large Language Models (LLMs) will become integrated to human decision-making processes, in many domains, from the mundane conversational agents, to more consequential medical assistance. Understanding human interaction with these systems, particularly in moral contexts, is therefore crucial.

Objective:

This research investigates two primary aspects: the ability of humans to detect whether judgments are generated by humans or LLMs (*detection*), and the degree to which they agree with these judgments (*alignment*).

Methods:

Our study involved a series of experiments beginning with a moral psychology test administered to 30 participants and two LLMs from the GPT-3 family. This initial phase aimed to assess how framing influences responses to moral dilemmas such as the trolley ('impersonal' dilemma) and bridge ('personal') scenarios. As previously reported (Greene et al. 2004), results indicated a weaker endorsement of utilitarian actions when dilemmas were framed 'personally' (i.e., they involve a physical involvement in the resolution of the dilemma), an effect that was even more pronounced in LLM responses.

Results:

Subsequent experiments focused on the evaluation of justifications for moral judgments, provided either by humans or LLMs in the initial experiment. These experiments (n=77 each) revealed that participants could generally identify the source of judgments above chance rate (~70%). We then turned to the analysis of the agreement and found that humans tended to agree more with AI judgements, specifically for complex 'personal' moral situations.

A critical finding of our research, however, concern the influence of participants' beliefs about the source of judgments on their agreement. Despite a general capability to detect AI-generated content, there was a significant dissociation between actual agreement with LLM judgments and agreement influenced by the presumed source. This bias suggests a complex interplay between recognition, belief, and acceptance of moral decisions from AI.

Finally, we analyzed the linguistic determinants of detection, and found that it can be significant reduced, but not eliminated, by modifying macroscopic linguistic features, such as length (humans produced shorter responses) and typos (humans made more errors).

Conclusions:

In conclusion, this study demonstrates the complexity underlying the processing of LLMs' judgements in moral and ethical decision-making domain, by highlighting far from perfect detection and inherent biases that could influence the deployment and trust in AI systems in sensitive applications.

P3-F-175 - You scratch my back and I'll scratch yours: Friendship closeness oppositely modulates hedonic responses during reciprocal benefit

Yang Xu [1], **Ping Wei** [1]

[1] Capital Normal University

Details

Objective: Friendships between two close friends are often viewed as reciprocal. Previous studies showed that closer friendship is associated with greater hedonic neural responses to vicarious rewards. However, much remains unknown about the neural mechanisms underlying the acquisition of reciprocal

rewards. Here, we used fMRI to investigate the neural mechanisms of gaining vicarious rewards for a friend and receiving reciprocal rewards from the friend and its association with friendship closeness (FC). Methods: Subjects (n=22) were asked to strive for rewards both for a friend and themselves, as well as observe the friend performing the same task. On each trial, subjects responded to a target following a cue indicating a potential reward for themselves, their friend, or neither as a control condition. Feedback indicating whether the reward was obtained was provided contingent on their response. Results: Behavioral results showed that reaction times (RTs) were both faster when gaining rewards for self and vicarious rewards, relative to the control condition. Neurally, univariate analysis revealed that, compared to negative feedback, stronger activation in the ventral striatum (VS) was associated with greater FC when gaining vicarious rewards. During observation, however, multivariate pattern analysis revealed that the classification accuracy of responses in VS was negatively associated with FC, when discriminating reciprocal rewards versus positive feedback of the control condition. More importantly, by using a neural signature of reward, we found that reward pattern expression of vicarious reward was positively associated with FC, while that of reciprocal reward was negatively associated with FC. Finally, using inter-subject representational analysis, we found that the more similar neural responses of the dorsal medial prefrontal cortex (dmPFC) to reciprocal rewards were, the more similar RT improvements of gaining vicarious rewards were. Notably, the connectivity between VS and dmPFC was increased during the cue phase of vicarious reward. Conclusions: To summarize, the hedonic neural responses were modulated in opposite ways by close friendship during reciprocal benefit, showing greater hedonic response to vicarious rewards, but lower hedonic response to reciprocal rewards. Concurrently, the reward system and the mentalizing network interacts in the transition between the roles of beneficiary and benefactor. These findings uncovered the neural mechanisms of reciprocal benefit between friends, suggesting its function of bonding close friendship.

P3-F-176 - Emotional correlates of antisocial behavior: peripheral nervous system markers and FaceReader study

Elizaveta Ivanova, **Oksana Zinchenko** [1], **Vladimir Kosonogov** [1], **Victoria Moiseeva** [1], **Anna Shestakova** [1], **Vasily Klucharev** [1], **Mario Martinez-Saito** [2]

[1] National Research University Higher School of Economics, [2] HSE University

Details

Objective

Antisocial behavior (ASB) is a broad concept that encompasses a range of behaviors that violate social norms, rules or laws, and disregards rules of conduct (Moffitt et al., 2015). Envy and schadenfreude are two forms of emotions associated with the antisocial behavior. Envy can be defined as a feeling of inferiority or discontentment with another person's possessions, abilities, or achievements. Schadenfreude, on the other hand, is the pleasure derived from observing another person's misfortune (Silver et al., 1978) associated with the emotion of happiness. Based on previous findings linking schadenfreude with antisocial behavior, we have decided to conduct a study to focus on facial expressions of the subjects engaged in modified economic games to evaluate the predictive power of facial expressions in free-riding behavior.

Methods

We employed the modified Ultimatum Game (Casal et al.,2011) to present our subjects with two conditions: «Punity» condition – where the responder can select the option of dividing the money between two subjects in such a way that only s/he will get the reward, while the proposer loses her/his budget, and standard «Ultimatum» condition – where, if the responder decides to reject the proposal, both participants will lose the reward. We have recorded the facial expressions of the responders at the outcome stage with the automated facial expression recognition software FaceReader. In addition, we have analysed the heart rate variability (HRV) and skin conductance response (SCR) of our subjects. The data was collected on 55 subjects.

Results

Repeated-Measures ANOVA analysis with two levels of factor *Game Condition* (Punity and Ultimatum) showed the significant difference in facial expressions recorded with FaceReader ($F(3,6238) = 7,036$, $p<0.001$) and HRV ($F(3, 6599) = 3,630$, $p = 0, 012$). Our findings suggest HRV was significantly higher in Punity condition and happy facial expressions were also more frequently present in Punity condition.

Conclusions

We aimed to investigate if particular combinations of facial expressions associated with schadenfreude can be linked to the acts of free-riding behavior. Our results showed that the automatic facial expression recognition algorithm can successfully detect the presence of happy emotional expression associated with the antisocial behavior in the economic game setting.

P3-F-177 - The Effect of Ecolabels and Visual Attention on Producer's Sustainable Decision-Making: A novel resource dilemma

Julian Kirschner [1], **Jan Hausfeld** [1], **Jan Engelmann** [1], **Natalie Lee** [1]

[1] University of Amsterdam

Details

Objective: Ecolabels, standards, and information provision are potential tools to shift producers and consumers towards sustainability. Yet, there is limited evidence from dynamic environments and their effects on producers. We address both by investigating the impact of ecolabels and information order on sustainable decision-making within a novel common resource paradigm, framed as producer decisions. In addition, we explore cognitive processes involved in the effect of ecolabels on attention and behaviour through eye tracking and by varying information order.

Methods: In a deforestation setting, participants withdraw trees (between 0 to 20 per round), from a common forest, with a steady regrowth rate of 40 trees per round (10), in return for individual profit. Conserving trees not only benefits the commons in the game, but remaining trees are also planted in the real world, creating a dynamic environment where environmental and individual outcomes are

incentivized. With this paradigm we study the effect of symbolic and monetarily incentivized ecolabels on behavior, by collecting data from 244 participants in the lab. To explore to cognitive processes involved and to understand in what way ecolabels bias attention and behaviour, we collected eye-tracking data from 96 participants and varied where information about economic and environmental impacts of their behaviour were presented.

Results: First, we find that purely symbolic and monetarily incentivized ecolabels reduce the initial (Eco M=10.65, NoEco M=13.36) and average amount of trees withdrawn, leading to more surviving common forests and environmentally (trees planted: +45%) as well as economically (trees cut: +18,4%) more efficient outcomes. Second, using eye-tracking and the randomization of information order we show that participants in the ecolabel conditions pay more attention to environmental outcomes. Through randomizing the order of information presented we can show a causal role of information salience, as information presented in the top left of the decision screen led to increased attention to the information, leading to participants that pay more attention to remaining resources, making more sustainable decisions.

Conclusion: Overall, we strongly emphasize the relevance of studying the effects of ecolabels and standards on producers, as shifts in their behaviour can have far reaching impacts on industries (Taufique et al., 2022). We conclude that ecolabels can be powerful coordination tools that facilitate collective shifts towards sustainable behaviour and shift attention of decision makers towards environmental outcomes.

P3-F-178 - Attention Discrimination and Performance Evaluation

Jan Hausfeld [1], **Yuki Takahashi** [2], **Boris Van Leeuwen** [2]

[1] University of Amsterdam, [2] Tilburg University

Details

Managers decide on the promotion/firing of workers. To accurately evaluate the worker's effort/performance, managers must pay attention to all workers. However, because people cannot pay attention to everyone, they must decide how to allocate their attention. This selective focus is likely affected by stereotypes, so called "attention discrimination". For example, a negatively stereotyped worker might receive less attention when the manager has to find the best worker, but receives more attention when the manager has to find the worst worker. This interplay between stereotypes and attention can lead to unjust promotion and firing decisions.

Objective: How do stereotypes affect managers' attention (eye-tracking) and evaluations of workers when not everyone can be tracked? Do these stereotypes prevail if live information is shown? How do self-reports of workers (overconfidence) affect the managers?

Methods: This project mimics how managers allocate attention in the workplace. Eye-tracked managers (N=424) saw 6 videos of workers performing a task at the same time and had to evaluate them. 105 workers were video-filmed doing manual tasks like creating metal planes or bracelets. All studies are pre-registered and sample sizes were based on simulated/estimated data for an expected medium effect size.

The managers' incentives varied such that it is sometimes more profitable to find the "best" worker, other times the "worst" worker.

Study 1: Managers (n=240) see 6 workers simultaneously performing tasks for 6 minutes and have to state the actual performance. We expect different allocation of attention dependent on the stereotypes of the task and the incentives ("attention discrimination" similar to Bartos et al. 2016), and are interested in how attention discrimination develops over time. In study 2 (n=60) and study 3 (n=124), we zoom in on the effect of only stereotypes and overconfident signals.

Results: We find that managers' attention and evaluation is affected by stereotypes. The negatively stereotyped workers receive less attention when the manager has to detect the "best" performers, and receive more attention and accurate evaluations when mangers find the "worst". Importantly, these differences in attention and evaluation are not driven by worker's true performance and certain workers suffer more. Showing continuous performance information does not overrule stereotypes.

Conclusions: Managers decide on who gets promoted/fired. To make these decisions, all workers should be tracked similarly. Since attention is scarce, "attention discrimination" affects these decisions leading to suboptimal decisions as "best"/'worst" performers are not accurately detected.

P3-F-179 - Neural mechanisms of social influence on taste preferences in mice: insights from automated, naturalistic behavioral testing

Mateusz Rycerz [1], **Maria Kalinowska** [1], **Fahmida Haque** [2], **Anna Madecka** [1], **Anjaly Yadav** [1], **Joanna Sadowska** [1], **Alicja Puścian** [1]

[1] Nencki Institute of Experimental Biology, [2] National Cancer Institute

Details

Objective: Understanding how social environments influence individual preferences is crucial for studying the neural underpinnings of decision-making. While individual taste preferences are well-documented, the dynamics of group conformity and the persistence of these preferences post-social influence are less explored. This study investigates how changing social contexts affect taste preferences in mice and explores the associated neural mechanisms.

Methods: We tested the voluntary, unshaped behavior of group-housed mice (up to 16) using the automated Eco-HAB assay, an open science system closely following murine ethology. Additionally, chemogenetics was employed to manipulate specific neuronal populations in the prelimbic cortex of transgenic mice. Mice were first tested for individual taste preferences in brief, 1-hour daily sessions over 9 days, where they had exclusive access to two flavors of candied milk with equal nutritional value. These sessions were the only times mice accessed the rewards, and preferences were recorded. The group housing environment was then modified to allow access to the same rewards in 1-hour sessions for 2 weeks, enabling mice to interact and influence each other's choices. Afterward, the design reverted to isolated testing to assess reversion to individual preferences.

Results: Initially, individual mice showed distinct preferences. During the group housing phase, these preferences shifted, with all mice conforming to the group's preferred taste. When returned to isolated testing, the mice reverted to their original preferences, indicating that social influence, while powerful, is transient. Chemogenetic activation of parvalbumin-expressing neurons in the prelimbic cortex selectively impaired sociability towards conspecifics, reducing social interactions and potentially impacting the transmission of social influence on individual preferences—a hypothesis currently under investigation.

Conclusions: Social environments can temporarily override individual preferences via group conformity. However, these changes are not permanent, as mice revert to their original choices when isolated. The Eco-HAB system facilitates precise, naturalistic observation of these dynamics, providing a robust platform for further research into the neural circuits involved in decision-making processes influenced by social contexts.

P3-G-180 - Individual differences in seeking and avoiding threatening information

Hayley Dorfman [1], **Elizabeth Phelps** [1, 2]

[1] Harvard University, [2] New York University

Details

People often have to decide whether to approach or avoid information about uncertain negative outcomes. For example, if you start coughing, do you choose to Google possible medical diagnoses, or do you procrastinate a visit to the doctor? Generally, people choose to avoid seeking out new information when in threatening contexts. For example, information avoidance is a common emotion regulation strategy often seen in people with anxiety. While it seems that threat largely biases people towards avoiding new information, some past work suggests that there are individual differences in information-related biases under threat. For example, anxiety can sometimes instead motivate people to seek out information about negative outcomes. While this past work suggests that people's tendency to choose to approach or avoid threatening information is highly variable, research has yet to determine how and why people differ on these information-seeking preferences.

Objective: This study characterizes individual variability in choices to seek or avoid information about the threat of negative outcomes. We present a novel theoretical and computational framework that can map individuals along dimensions of information-seeking and information-avoidance by considering the arbitration between the desire to minimize negativity and the desire to minimize uncertainty.

Methods: To test this framework, we developed a unique behavioral task that assesses information seeking/avoidance behavior by decomposing components of information into information likelihood and information noise. In order to assess choice behavior about a realistically stressful scenario, participants will undergo a threat of shock procedure where they are asked to choose to receive or avoid information about whether they are in a "shock" state or a "safe" state. Changes in pupil dilation will be used as a confirmatory measure of autonomic arousal during this manipulation. To investigate whether choice behavior is associated with individual differences, participants will also complete a series of questionnaires assessing their levels of anxiety, depression, intolerance of uncertainty, and intolerance of negativity.

The results of this work will help elucidate the precise circumstances under which people seek information about threat, avoid information about threat, and arbitrate between the two strategies.

P3-G-181 - Reference-dependence and mood: anhedonia and feeling down are associated with divergent reference points

Agnieszka Tymula [1], **Paul Glimcher** [2], **Nicholas Glozier** [1], **Alexander Svenson** [1]

[1] University of Sydney, [2] New York University

Details

Objective: Mood disorders such as major depressive disorder manifest as different yet overlapping symptom clusters across individuals, with heterogeneity in aetiological factors that can complicate diagnosis and treatment. Building on past research on reference dependence and mental health, we develop and test the hypothesis that two key diagnostic symptoms of depression: (1) anhedonia and (2) feeling down, depressed, or hopeless; are associated with miscalibrated and diverging reference points. **Method**: We recruited a representative sample of 443 participants on Prolific. We measured participant's reference points by both estimating them from a lottery choice task and through self-reports. The choice task involved deciding whether one preferred a lottery with two possible (probabilistic) outcomes or a sure outcome. Using Expected Subjective Value Theory (Glimcher and Tymula, 2023), we estimated participants' reference points based on their decisions in the lottery task. We measured self-reported reference points as expectations of task earnings and income expectations. We also asked subjects to complete Patient Health Questionnaire-9 (PHQ-9) which assesses the severity of multiple depressive symptoms. We applied maximum likelihood estimation and regressions to investigate the association between depressive symptoms and reference points. **Results**: Anhedonia (PHQ item 1) was associated with higher reference points while feeling down/depressed/hopeless (PHQ item 2) was associated with lower reference points. Our estimates indicate a person reporting anhedonia nearly every day (highest severity) would have a reference point 75% higher, while the reference point is nearly halved for those reporting feeling down nearly every day. While the statistical strength varied, the same directional pattern was found across all our measures of the reference point. The income expected in one year by someone reporting persistent anhedonia only compared with feeling down only was estimated to diverge by over 40,000 USD. Moreover, risk tolerance and response times diverged in expected ways based on reference dependent value functions: high severity anhedonia increased the frequency of risky choice by 16%, while high severity feeling down decreased it by 11%. Preliminary data from clinically depressed treatment-resistant patients indicate reference dependence may covary with symptoms severity across two time-points, supporting our cross-sectional analysis. **Conclusion**: Our results evidence that reference-dependence may be useful for understanding the aetiology of mood-disorders, including differences across symptom clusters, holding promise for improved diagnosis and treatment.

P3-G-182 - Predicting children's current and future mental health by their brain functional connectivity

Mirjam Habegger [1], **Nora Raschle** [1], **Todd Hare** [1, 2]

[1] University of Zurich, [2] University of Zürich

Details

Objective: Childrens' functional connectomes are associated with their mental health. Here, we test how the state of brain connectivity predicts aspects of current and future mental health.

Methods: In a random subsample of 1000 children (aged 9-10y) from the *Adolescent Cognitive Brain Development* (ABCD) study, we run cross-validated linear regressions to predict two scales of the child behavior checklist (i.e., *stress* and *internalizing problems*) using functional connectivity. We applied dimensionality reduction via a principal component (PC) analysis of resting-state functional connectivity within and between cortical networks and between cortical and subcortical regions. We used the first250 PCs in the regression models to explain 1) stress and 2) internalizing scales at the current state (9-10 years) and 2 years later (11-12 years).

Results: Children's brain functional connectivity at age 9-10y explained ~1% of children's current stress state and internalizing problems ($p = 0.005$ and 0.010, respectively), and ~0.5% of stress and ~1% of internalizing problems 2 years later at age 11-12y ($p = .018$ and 0.007).

The loadings of significant PC's revealed that cortical-subcortical connections were more often important for predicting mental health than connections within the cortex. E.g., fronto-parietal network connections to the left or right amygdala were significant for current and future stress and internalizing problems. For the stress scale, connections between the amygdala and the ventral attention network were also important at both time points. For internalizing scale, connections between the amygdala and the default mode network and the cingulo-opercular network were additionally relevant for predictions. However, some cortical-cortical connections (e.g., cingulo-parietal network and attention network) were also important across health measures and assessment waves.

Conclusion: We found significant links between functional connectivity and children's mental health both concurrently and 2 years later. Unlike many previous studies, we examined both cortical-cortical and cortical-subcortical connections. We found that the latter were particularly important for predicting mental health. Connections between networks associated with crucial socioemotional skills, such as emotion regulation, were significant across measures and assessment waves. Understanding more about the associations between brain connections and mental health, especially the brain's predictiveness for future mental health states, may provide insights into mental health development and guide early interventions and preventions.

P3-G-183 - In-utero trauma during World War II affects brain development and life outcomes

Gökhan Aydogan [1], **Sepp Kollmorgen** [1], **Valerio Mante** [1], **Birgit Kleim** [1], **Gideon Nave** [2], **Christian Ruff** [1]

[1] University of Zürich, [2] University of Pennsylvania

Details

Armed conflicts are a regrettable yet persistent part of human life, and their recent resurgence has highlighted the question of how such disruptive events influence the neural development of exposed children and affect their long-term life trajectories. Previous research indicates that traumatic events in early life adversely affect brain structure during critical stages of neural development (Teicher et al., 2018) and are statistically linked to depression and anxiety disorders later in life (Klika et al., 2020). However, the presence of numerous confounding factors such as socio-economic status and genetic makeup, and the small sample sizes of these studies, leaves it unclear whether the observed associations actually indicate that early life trauma causally affects both brain development and long-term life outcomes.

Here, we utilize a natural experiment to investigate in a large sample (N=37,856) how in-utero exposure to traumatic events during various phases of pregnancy impacts adult brain structure and life outcomes. To this end, we combine the largest collection of publicly available brain images, the UK Biobank (UKB), with data from the British National Archive that includes meticulously recorded bombing locations resulting from German air raids during World War II. By linking birth dates and locations with timing and location of bomb explosions, we were able to evaluate the causal nature of the relationship between brain structure, life outcomes, and in-utero exposure to air raids, while systematically controlling for genetic and other confounding environmental factors such as socio-economic status or maternal smoking.

We observed a significant reduction in Grey Matter Volume (GMV) associated with exposure to bombings, specifically for those occurring within a small radius (<2km) of the birth location during the last trimester of gestation ($p<10^{-8}$), with notable reductions primarily in prefrontal and cerebellar brain regions. In line with GMV alterations, we find substantial effects on various white matter fiber tracts related to important cognitive abilities (e.g. fluidic IQ), such as the corpus callosum, superior fronto-occipital fasciculus and the posterior thalamic radiation ($p<10^{-4}$). Mediation analysis reveals that this reduction in GMV is linked to lower IQ and higher BMI ($p<10^{-5}$) measured later in life, while controlling for the genetic predispositions for these phenotypes.

Our results provide well-powered and robust evidence that traumatic events during pregnancy have an enduring impact on human neural development and associated life outcomes, and underscore the importance of safeguarding also unborn life during armed conflicts or similarly stressful events.

P3-H-184 - Rats and Humans are Susceptible to Context Biases in Reinforcement Learning

Magdalena Soukupova [1], **Lachlan Ferguson** [2], **Stefano Palminteri** [3], **Shauna Parkes** [2]

[1] Ecole Normale Superieure, [2] University of Bordeaux, [3] École Normale Supérieure

Details

Objective

It has been shown that humans as well as other animals, including rats, are capable of transitive inference. That is, if they learn that A>B and B>C, they are also able to infer that A>C. However, when options are

learnt in fixed stable pairs (contexts), humans are often biased by the original learning context, preferring options that were good in the original context over options with higher expected value that were originally bad. In this study, we investigated whether such a context bias also occurs in rats.

Methods

To study this we conducted two sets of experiments, one for each species. Each consisted of a learning task followed by a transfer test. In the learning task, subjects (16 rats; 24 humans) encountered options with different outcome probabilities presented in fixed stable contexts, namely AB [100%; 62.5%] and CD [37.5%; 0%]. To ensure that all subjects had the same knowledge of all options, humans received full feedback after each choice (the outcome of both the chosen and the unchosen option was shown), while the learning task for rats was structured as a mixture of rewarded forced-choice trials and unrewarded free-choice trials. In the transfer test, subjects had to choose between previously learnt options that were repeatedly presented in all possible pairwise combinations (i.e. AB, AC, AD, BC, BD, CD).

We then used mixed model logistic regression to calculate the probability of choosing the option with the higher expected value in each of the possible transfer test pairs. The BC comparison was of particular interest, as it was the only comparison in which the option with a higher expected value (B) had a lower relative value and vice versa. Thus, a strong preference for C would indicate context-dependent encoding, whereas a preference for B would indicate absolute, i.e. context-independent, encoding.

Results

Both rats and humans strongly preferred option C over option B in the transfer test (rats: P(B) = 0.356, 95% CI [0.280, 0.440], z=-3.29, p=0.001; humans: P(B) = 0.292, 95% CI [0.181, 434], z=-2.794, p=0.005), suggesting that relative values likely played a preponderant role in their decision-making.

Conclusions

Both species appeared to encode values in a context-dependent manner. It is therefore likely that the observed context bias stems from common cognitive mechanisms shared between the species rather than factors specific to human evolution or culture.

P3-H-186 - Value-Based Decisions Across Multiple Goals

Ali Shiravand [1], **Yumeya Yamamori** [2], **William Bhot** [2], **Stefano Palminteri** [3], **Benedetto De Martino** [2]

[1] École Normale Supérieure-PSL University, [2] University College London, [3] École Normale Supérieure

Details

People often follow multiple goals when making decisions in their daily lives. For instance, when choosing a career, individuals might consider factors such as financial stability, personal fulfillment, and work-life balance all at once. At the same time, commuting time to the office plays a less important role when choosing between job offers. Studies have shown that people can use previously learned decision-making strategies when confronted with a new goal, although the underlying mechanism of how they can use their learned values to fulfill multiple goals has been underexplored. In this study, we developed two novel multi-goal reinforcement learning (RL) tasks that we tested on human participants (Experiment 1 in-person N=30 and Experiment 2 online N=105). We developed and tested two competing RL models that can explain learning across multiple goals. Our modeling results indicate that a group of participants combined information relative to their goal into a single value during learning, however, the rest learned the value associated with each goal before making decisions and integrated the learned values at the time of decision. Although the latter approach looks computationally more expensive, it allows the second group to adapt faster while facing a new multi-goal choice problem (two-sided Mann-Whitney test, $U=417.5$, $P<0.01$, fitted learning rate$\in(0.05,0.95)$ and temperature$\in(1,100)$, respectively). In addition, we show that people who were more confident in their choices were also more adaptive in their new multi-goal decisions (OLS Regression, $F=47.9$, $P<0.001$). This study provides new insights into underlying computational mechanisms of how people use multidimensional values to pursue multiple goals.

P3-H-187 - What is the value of affect during self-belief formation?

Nora Czekalla [1], **Alexander Schröder** [1], **Annalina Mayer** [1], **David Stolz** [1], **Frieder Paulus** [1], **Sören Krach** [1], **Laura Müller-Pinzler** [1]

[1] University of Lübeck

Details

Self-beliefs and beliefs about the world are formed and constantly updated in the face of social feedback that deviates from expectations, also called prediction errors. Considering the subjective value of certain beliefs may explain biases in belief updating, especially for self-beliefs. Here, the current affective state during belief updating is particularly important.

In two consecutive functional MRI studies, we examined neural activity and pupillometry associated with prediction error-based learning about oneself and another person as well as the current affective state. We computationally modeled the trial-by-trial formation of self-beliefs and beliefs about another person and collected affect ratings at regular intervals. In the first study, we measured a healthy sample (n = 39 MRI, n =30 behavior only, Müller-Pinzler et al., 2022); in the second study, we compared a healthy replication sample (n = 32) to individuals diagnosed with depression (n =35) as a model for altered affective conditions (Czekalla et al., 2024). We hypothesized that more positive and negative affective states are associated with more positively and negatively biased formation of self-specific beliefs, respectively, and altered neural tracking of self-related prediction errors.

The results of both studies showed that biases in self-belief updating were associated with the intensity of experienced affect during the task. Neural activity within the anterior insula, amygdala, ventral tegmental area, and medial prefrontal cortex during the experience of prediction errors was linked to biased updating as well as affect. Specifically, participants who tended to update their self-related beliefs

more negatively had stronger neural and pupil responses to negative prediction errors and experienced more embarrassment and less pride. In the second study, a higher depressive symptom burden was associated with forming more negative self-beliefs and more positive beliefs about others.

Our results provide evidence for neural and computational mechanisms that integrate affect into the process of belief formation.

References:

Czekalla, N., Schröder, A., Mayer, A. V., Stierand, J., Stolz, D. S., Kube, T., Wilhelm-Groch, I., Klein, J. P., Paulus, F. M., Krach, S., & Müller-Pinzler, L. (2024). Neurocomputational Mechanisms Underlying Maladaptive Self-Belief Formation in Depression. In *bioRxiv* (p. 2024.05.09.593087).

Müller-Pinzler, L., Czekalla, N., Mayer, A. V., Schröder, A., Stolz, D. S., Paulus, F. M., & Krach, S. (2022). Neurocomputational mechanisms of affected beliefs. *Communications Biology*, *5*(1), 1241.

P3-H-188 - Neurocomputational mechanisms of affected beliefs

Sören Krach [1], **Alexander Schröder** [1], **David Stolz** [1], **Annalina Mayer** [1], **Nora Czekalla** [1], **Frieder Paulus** [1], **Laura Müller-Pinzler** [1]

[1] University of Lübeck

Details

Self-beliefs, such as beliefs about our abilities, attractiveness, or personality, are under constant (re)evaluation depending on the feedback and information we receive from our surrounding world. However, feedback processing is not a passive process during which information is picked up objectively, rather the idea prevails that belief formation is essentially biased and shaped by affective and motivational processes. In several studies, using the Learning-of-own-performance task (Müller-Pinzler et al., 2019, 2022), we approach the question of how humans arrive at these self-beliefs in the first place (study 1; N=60) and, once established, how these self-beliefs are revised in the face of conflicting evidence (study 2; N=102). Using computational modeling, functional neuroimaging, and psychophysiological data, we show that the formation of self-beliefs is biased towards negative information and this bias is associated with the experience of affective states during belief formation. The results further suggest that individuals who update their beliefs more negatively and experience stronger negative affect, process negative information more intensely than positive information as indicated by increased pupil dilation and neural activation within the insula and amygdala (study 3; N=60). Finally, in a clinical sample of persons diagnosed with major depression (MDD) and healthy controls (HC), we replicate the negativity bias by showing that both groups have similar patterns of negatively biased belief formation. Further, in the insula negatively biased updating was accompanied by stronger tracking of negative, but not positive, prediction errors with increasing symptom severity (study 4; MDD=35; HC=32). Our findings provide support from healthy and clinical samples for the

overall rationale of the formation of affected beliefs, that is, the notion that beliefs are fundamentally shaped by motivational biases as well as affective experiences during feedback processing.

P3-H-189 - Cost-benefit tradeoffs in integrating information in multiple-feedback Reinforcement Learning

Maël Lebreton [1], **Antonios Nasioulas** [1]

[1] Paris School of Economics

Details

While an increased quantity of information can improve learning and decision-making, integrating more information is also costly. Consequently, cost-benefit tradeoffs emerge when one is faced with an increased quantity of information. This is a highly neglected aspect in the Reinforcement Learning (hence, RL) literature, where, in the typical set-up, at each round, one receives one feedback sample for the chosen option (which is also one's reward) and potentially one feedback sample for the unchosen option (in case of complete feedback). In a series of three behavioral experiments (total N = 300), we explore an innovative RL framework, in which one receives multiple feedback samples associated with each option. The objective of our study is to investigate how participants treat this extra information, and how their learning performance and their confidence are affected by it. Our results show that, as expected, more information (more feedback samples) leads to better learning (N = 2x100, P = e-06, d = .44). Despite the effect on learning and even though there is also high meta-cognitive sensitivity (confidence and performance are positively related; linear regression between accuracy and confidence quantile gives an estimated coefficient of 0.028, P = e-93), confidence seems to be only marginally affected by the number of feedback samples (N = 2x100, P = .02, d = .19). The increased duration of looking at the samples when more samples are present (N= 2x100, P = e-09, d = .33) and the absence of any difference at the time of choice (Reaction Times; N = 2x100, P = .38) are compatible with a standard Q-learning model -using cached Q-values at the choice stage. This new rich framework allows us to disentangle the reward dimension from the informational dimension. We consider a model that tracks both Q-values (reward prediction error) and I-values (information prediction error). Model comparison between this, the pure-reward (Q-values only) and the pure-info (I-values only) model greatly favors the pure-info model -which is the winning one for about 80% of the participants. Finally, exploratory analyses reveal that there seems to be a context effect (the increased accuracy being associated with the high-information condition within a context, rather than with the absolute number of samples given) and a limited cognitive budget within sessions (sessions with a higher aggregate number of samples did not lead to better performance).

Our results suggest that while more information improves performance in RL, complex cost-benefit tradeoffs in integrating this information emerge and participants seem to strategically allocate their limited cognitive resources.

P3-I-190 - generalized grid-like coding of neuroeconomic value

Mark Orloff [1], **Seongmin Park** [2], **Jake Blumwald** [1], **Philippe Domenech** [3], **Erie Boorman** [1]

[1] University of California, Davis, [2] National Center for Scientific Research, [3] Institut national de la santé et de la recherche médicale

Details

When making choices, humans often need to be able to choose between items with different characteristics (e.g., a bottle of water vs. a bag of chips). The brain valuation system solves this problem with a 'common currency' (Levy & Glimcher 2012). This value coding has been identified in dorsal anterior cingulate cortex (dACC; Kennerley et al., 2014), orbitofrontal cortex (Kennerley et al., 2014; Padoa-Schioppa et al., 2006, 2008), and ventromedial prefrontal cortex (vmPFC; Chib et al., 2009; Lebreton et al., 2009). Our recent work shows that in addition to this brain valuation system, a grid-like representation, a key component of the cognitive mapping system, emerges in a risky decision-making task in entorhinal cortex (EC) and mPFC (Orloff et al., 2023, Neuroeconomics). In this task, individuals made binary choices between shapes, which are drawn from two sets (circles and triangles) that each vary along two continuous dimensions corresponding to reward magnitude and probability. Crucially, participants never made choices *between* shape sets (i.e., circle vs triangle). Prior animal work suggests that when two distinct contexts are merged (via the removal of a partition separating two learned rooms), their grid maps become fused together (Wernle et al., 2018). Here, we hypothesize that this will occur when individuals are required to make choices *between* shape sets after developing separate grid-like representations for each context. We collected data from participants across two scanning sessions. In the first session (Orloff et al., 2023, Neuroeconomics), participants only made choices *within* shape sets (task one, N = 35). In session two, participants made choices *within* and *between* shape sets (task two, N = 32). We test how 1) the brain valuation system and 2) grid-like representations (when making *between* set choices) change as separately learned systems are combined in task two. In both task one and task two the brain valuation system (vmPFC, dACC, and bilateral LPFC) significantly encoded the value difference. However, there were no significant differences between tasks. We observed significant grid-like representation in right EC and bilateral mPFC in task one. In task two, we observed significant grid-like representation in only left mPFC. Representations in right EC and right mPFC were significantly greater in task one than task two. However, bilateral rostral (r)mPFC had significant grid-like representation in task two but not task one. These data replicate previous work that the brain valuation system is a generalized value system. Further, we show that the rmPFC may be responsible for merging previously distinct grid-like representations.

P3-I-191 - Tracking unsignalled effort in volatile environments

Rong Bi [1], **Patricia Lockwood** [2], **Lilian Weber** [1], **Miriam Klein-Flügge** [1]

[1] University of Oxford, [2] University of Birmingham

Details

Humans flexibly adjust their behaviour to features of the environment, such as its volatility. Previous work has mostly focused on volatility in the reward domain. However, it remains unclear whether people are able to track and adjust their behaviour to the volatility of other environmental features, such as changes

in effort demands. Indeed, in naturalistic environments (e.g., cycling in more or less windy conditions), effort is not usually signalled, but can be predicted from experience.

We designed two experiments in which human participants tracked the varying physical effort of one (N=34) or two options (N=28), respectively, over time. Effort was implemented using hand-held dynamometers. In both studies, we manipulated the volatility of effort changes. In the second experiment, we also manipulated the average effort and acquired 7T-fMRI to investigate neural signatures of effort volatility learning.

In the first experiment (one option with varying effort), participants were able to learn about and predict effort requirements in the absence of explicit information about effort. Moreover, people's effort adjustments reflected both the noise ($F_{(1, 33)}$ = 35.02, p < 0.001, η^2_p = 0.507) and the volatility ($F_{(1, 33)}$ = 10.77, p = 0.002, η^2_p = 0.241) of the environment. In the second experiment, participants were more likely to stay with their current choice in stable compared to volatile blocks when the overall effort was low ($F_{(1, 27)}$ = 9.74, p = 0.004, η^2_p = 0.265), and higher levels of required efforts led to an increased frequency of choice switches ($F_{(1, 27)}$ = 14.99, p < 0.001, η^2_p = 0.357). Preliminary analyses of 7T-fMRI data suggest uncertainty signals in median raphe and average effort signals in the habenula and medial prefrontal cortex (mPFC).

Our findings show that effort learning reflects important environmental features, such as volatility and average costs. In ongoing work, we are examining the neural signatures underlying such flexible behavioural adjustments in the effort domain.

P3-I-192 - Detecting regime shifts in financial investment decisions

Ai-Lin Chen [1], **Mu-Chen Wang** [1], **Shih-Wei Wu** [1]

[1] National Yang Ming Chiao Tung University

Details

The world is in a constant state of change. Regime shifts represent a specific form of change wherein the world's state abruptly transitions from one to another. However, detecting a regime change is often challenging due to the noisy nature of the signals we receive and their susceptibility to environmental volatility. It is important to note that regime shifts have the potential to lead to either favorable or unfavorable outcomes. To investigate how potential shifts to desirable regimes or undesirable ones can impact change detection, we designed a regime-shift detection task mimicking the financial markets. In each trial, participants would begin in either a good or bad market and receive signals of upward or downward trends at every period. At any point during a trial, the market may switch from one regime to another. The subjects' task was to provide their estimates regarding whether the regime had shifted after observing each signal (10 sequential signals per trial), and subsequently, they were required to make a decision on whether to invest in the market based on the signals. We manipulated the system parameters by varying the transition probability and signal diagnosticity. Our findings revealed that subjects (n=33) exhibited system neglect in change detection: Overreactions to change were observed when the signals were noisy but the environments were stable. In contrast, underreactions were observed when the signals were precise but the environments were unstable. When comparing the

impact of market-shift directions, we found no significant difference in subjects' sensitivity to system parameters. However, individual differences in bad-to-good direction were significantly smaller than good-to-bad direction. Moreover, system parameters also influenced investment decisions. In a good market, participants were less willing to invest in the market when environments were more unstable or when they received noisy signals. Conversely, in a bad market, participants were more willing to invest in the market when environments were unstable. Additionally, signal diagnosticity had a greater impact on investment decisions when participants were in more stable environments. Together, our study replicated previous findings on system neglect and elucidates the impact of good/bad market transition and system parameters on investment decisions.

P3-I-193 - Computational Properties of Human Reasoning as Emergent Dynamical Features of Time-Optimal Inference

Francesco Trapani [1], **Carlos Stein** [1], **Daniel Mcnamee** [1]

[1] Champalimaud Foundation

Details

Humans are endowed with cognitive skills to make optimal choices and solve problems in complex environments. A full understanding of the cognitive processes involved in human reasoning should account for all of the three levels theorized by Marr (Marr, 2010): a computational theory, the algorithms and representations implementing such theory, and the neural basis that embodies it. Pioneering works in cognitive science and symbolic AI attempted to explain the computational theory of human cognition in terms of knowledge search processes (Newell & Simon, 1972). A set of key features and paradigms were identified - among the others: means-end analysis, hierarchical abstraction, least commitment principle, allocentric vs egocentric planning - that define and explain cognitive processes (Morris & Ward, 2004). Nevertheless, these studies were confined to the computational level and did not try to make connections with representational or biological bases. To our knowledge, a comprehensive theory that normatively establishes and explains how such computational principles arise from specific algorithmic and representational bases is missing.

Here we present and test a dynamical systems framework for flexible inference that normatively explains such computational paradigms. We express our theory in the domain of discrete Markov decision processes. Our algorithm relies on a compressed non-generative model of the world and updates its policy through exact inference with gradient ascent. As opposed to traditional approaches that rely on simulation-based inference or "rollouts" similar to search processes (Newell & Simon, 1972), our model is anytime-optimal, as it leverages the full structure of the task space at each policy update, providing an optimal trade-off between time and value returned. As we show, these choices alone at the algorithmic level result in the emergence of well-established computational properties of higher cognitive processing within the associated policy optimization dynamics. For example, our model dynamically exhibits hierarchical abstraction, as it converges to optimal solutions by automatically prioritizing policy updates on subgoal (or bottleneck) states. Furthermore, we propose an inductive bias mechanism within our framework which modulates the situational focus (i.e. the trade-off between allocentric and egocentric planning) of the planning process. This is naturally meta-controlled via heterogeneous weights on the

gradient vector. Taken together, we provide a unified normative quantitative framework for many of the key features of expert human reasoning as proposed in the era of symbolic AI that hitherto have remained unexplained.

P3-I-194 - Linking Gender Differences in Reference Points and Labor Market Outcomes

Lexin Liang [1], **Paul Glimcher** [1]

[1] New York University

Details

Objective: Despite significant female education advances, we continue to see gender disparities in income and labor market outcomes. Prior studies suggested these disparities may be driven by lower expectation among women, leading them to accept less favourable outcomes. This study aims to investigate gender differences in reference point, an expectation-based benchmark against which a decision maker evaluates options. we hypothesize that women have lower reference points than men, which predisposes them to engage in tasks with limited individual benefits.

Method: We employed a modified version of Babcock et al.'s (2017) investment game. In the original game, subjects are assigned to groups of three and presented with an investment opportunity that only one person can undertake. Subjects are informed that investing results in a lower payoff ($1.25) compared to non-investors ($2). However, if no group member invests within the time limit, all members earn the lowest payoff ($1). This effectively creates a tension where group members wait for others to invest whenever possible. This creates tension, as group members wait for others to invest whenever possible.

In our study, we introduced two manipulations to enhance the generalisability of the findings and to make the task more appropriate for a future fMRI experiment. First, we varied the investor payoffs to manipulate the incentive associated with investing. Second, we increased the group size to five people and varied the proportion of women in the group. We will assess how subjects' investment rates vary as a function of investor payoff and the proportion of women in the group. Classic economic models will also be applied to quantify the gender gap in reference points.

We aim to recruit a total of 200 participants (aged 18-65) in total, with 20 participants completing 12-18 rounds of the game per session.

Results: Consistent with Babcock et al.'s (2017) findings, our preliminary data (N=56) suggests that women consistently over-invest compared to men across a range of different investor payoffs. In mixed-gender groups, we also found that subjects invested more often with higher proportions of men in the group, and less often with high proportions of women in the group.

Conclusions: Our findings will elucidate how social and economic inequality influences women's decision-making and guide the development of strategies to reduce gender-based disparities. Future research will investigate neural underpinnings expectation signals that drive gender differences in decision-making.

P3-I-195 - Do inherent preferences interfere with associative learning of food values?

Alexandra Rich [1], **Sohum Kapadia** [1], **Ohad Dan** [1], **Ifat Levy** [2]

[1] Yale University, [2] Yale School of Medicine

Details

Background: Cognitive and behavioral flexibility is crucial for adapting to changing environments. Studies probing this often use monetary rewards; however, food valuation is evolutionary and reflects strong, heterogeneous preferences beyond nutritional content. These inherent – rather than learned or calorically assumed – values motivate food choices in healthy individuals and may critically drive disordered eating behavior. Prior work suggests that such values can be changed, yet very little is known about the flexibility of this process.

Objective: Characterize the behavioral mechanism implicit in updating inherent food values.

Methods: We designed a novel probabilistic reversal-learning task in which participants (n=200) predict associations between personally salient food stimuli and point outcomes. Inherent subjective values were estimated from pre-task ratings of 65 food items, from which 4 non-neutral items were randomly selected for each participant. In "Aligned" blocks, participants' subjective value ratings align with outcomes; in "Unaligned" blocks, this relationship is reversed, requiring flexible updating of food-value associations. We used model-free and model-based approaches to probe differences in food-value learning, as reflected in participants' outcome predictions, between Aligned and Unaligned conditions.

Results: Preliminary data (n=19) indicate that participants exhibit lower average prediction accuracy on Unaligned than Aligned trials (p=.007), suggesting the significance of inherent food value in learning flexibility. We performed model simulation and fitting for Rescorla-Wagner models with (1) a fixed learning rate (α, ranging from 0 to 1) and (2) separate Aligned and Unaligned learning rates. In the complete sample (n=200), we will perform model comparison to uncover latent variables that may best describe learning deficits observed in the Unaligned condition.

Conclusions: Preliminary results suggest the general population has significant difficulty learning associations that challenge food-value priors. In addition to motivating food choices in healthy individuals, these inherent values may critically drive disordered eating. Obesity is a major health challenge and binge eating characterizes a particularly harmful phenotype of obesity. Binge eating is associated with impaired cognitive flexibility and goal-directed control. Dietary interventions, which rely on flexibly updating food-value associations, are less effective in those who binge eat. As such, characterizing flexible inherent food value updating can inform integrated and personalized dietary interventions for obesity and binge eating.

PS.03.02, P3-I-196 - Bounded rationality in rats is accounted for by value normalization

Margarida Pexirra [1], **Jeffrey Erlich** [2]

[1] University College London, [2] NYU Shanghai

Details

Models of choice behavior typically rely on fundamental assumptions of rationality based on well-defined economic axioms. The Independence of Irrelevant Alternatives (IIA) axiom is one of these premises of rationality, stating that the relative preference between two options should be unaffected when extra options are added to the choice set. However, studies across different species have demonstrated violations of the IIA axiom, in a way that is consistent with the predictions of divisive normalization. To investigate the neural mechanisms that produce violations of the IIA axiom under a value normalization framework, we developed a task where rats make decisions between six options that differ in reward value. A trial was initiated when rats enter their snouts in a central nose port, triggering lights that indicate which of the six options were available for choice on a given trial. Forty-one different trial types were presented in an interleaved manner during a session: six where only one port was available, fifteen pairs, and twenty triplets. After a variable amount of time, a sound go-cue was played, releasing animals from fixation and allowing them to make a choice towards any of the illuminated available ports. Once selected one of the available ports, a volume was sampled from the reward distribution associated with the the chosen port and delivered via a dedicated reward port. This task structure, featuring the same option in multiple choice contexts, allowed us to test how the presence and value of a third (unchosen) option influenced choice ratios - a key metric of violations of the IIA axiom. Animals learned to choose the highest value port available on a trial-by-trial basis: developing stable and transitive preferences between two ports. Interestingly, in trials with a third irrelevant option (distractor), animals almost never chose the distractor, but the target choice ratio between the top two options was influenced in a way that was consistent with value normalization. Having established a framework for studying value normalization in rats, we are now collecting neurophysiological data from regions involved in economic decision-making, such as the orbitofrontal cortex and the frontal orienting field and planning causal perturbation studies to investigate how and where value normalization is implemented.

P3-I-197 - Navigating uncertainty: the role of confidence in reinforcement learning

Constance Destais [1], **Nahuel Salem-Garcia** [2], **Stefano Palminteri** [2], **Maël Lebreton** [1]

[1] Paris School of Economics, [2] École Normale Supérieure

Details

Previous studies show that humans can adapt their learning rate to uncertainty in the environment, as characterised by the objective volatility. Here, we test the alternative, original hypothesis that learning rates are modulated by the *subjective* estimate of choice accuracy: *confidence*. We leverage recent

findings (the valence-induced confidence bias - Lebreton et al., 2019) and a unique experimental paradigm (two-armed bandit with a valence manipulation and confidence elicitation) to disentangle the role of objective uncertainty and subjective confidence in modulating learning rates.

200 participants, recruited on Prolific, completed our task. A between-subject manipulation of outcome information yielded a key experimental design (partial feedback, n=100) that maximises the effect of valence (gain vs. loss outcomes) on confidence. In this design, valence affected participants' confidence ($t(99)=9.96$, $p<.001$, Cohen's $d=.99$) but not their choice accuracy ($t(99)=1.24$, $p=.22$, Cohen's $d=0.12$). Crucially, participants' propensity to switch choices was also affected by valence ($t(99)=5.97$, $p<.001$, Cohen's $d=.60$). This result cannot be accounted for by the effect of objective uncertainty on learning and suggests that instead, subjective confidence modulates the learning process.

Expanding upon existing reinforcement learning (RL) frameworks, we introduce a new model with dynamic learning rates that are modulated by confidence. We fit the parameters of this new model and of standard RL models to each participant's trial-level choices. The model with confidence-dependent learning rates provided a better account for participants' choices than standard RL models (protected exceedance probability>.99). Finally, our model-based analyses build on RL models that distinguish learning from confirmatory and disconfirmatory information, so we compare confirmatory and disconfirmatory learning parameters in the new model to see whether confidence modulated them symmetrically. The significant difference between confirmatory and disconfirmatory learning parameters ($t(199)=5.07$, $p<.001$, Cohen's $d=.36$) suggests that high confidence exacerbates the confirmation bias in learning.

Our model-free and model-based analyses both provide evidence for the effect of confidence on reward learning. Confidence-modulated learning rates could emerge as a parsimonious mechanism of how learning adapts to varying degrees of uncertainty. Our model-based analyses, using a novel confidence-dependent RL model, also suggest that subjective confidence in choice accuracy can shape learning biases.

P3-I-198 - It's sad but true: Emotions and trustworthiness of information

Marja-Liisa Halko [1], **Juho Halonen** [1], **Marita Laukkanen** [2], **Henri Nyberg** [3], **Mikko Salmela** [1]

[1] University of Helsinki, [2] VATT Institute for Economic Research, [3] University of Turku

Details

Objective: In our modern era, marked by the prevalence of social media and intricate economic and political challenges, the influence of information has become increasingly polarized. Misinformation, strategically crafted to manipulate opinions and voting decisions based on biased or inaccurate premises, has notably impacted election results in the US and Europe. Faced with diverse information sources, individuals must decide whom to trust. This raises critical questions: How do individuals navigate decisions of trust in information?

Methods: We conducted a behavioral experiment with 62 participants, using physiological measurements, including galvanic skin response (GSR), heart rate (HR), and facial expression analysis, to assess the emotional impact on information trustworthiness. Participants evaluated short information

excerpts, true or false, drawn from various sources such as news media and research articles. They rated the trustworthiness of each excerpt, related to polarizing topics like economic policy, immigration, and climate change. We utilized physiological measures, along with machine learning methods, to predict participants' trustworthiness ratings, hypothesizing that physiological data, alongside political ideologies, will enhance the predictive accuracy of the participants' trustworthiness evaluations.

Results: We uncovered conformism, or a political self-serving bias, evidenced by the significant positive correlation between participants' conservatism scores and the perceived trustworthiness of pro-conservative statements. In addition, we found that immediate emotional reactions, specifically an increase in the expression of disgust and sadness and a decrease in the expression of fear, were significantly associated with trustworthiness ratings among the most conservative participants, suggesting that ideological leanings might intensify the impact of certain emotional responses on the assessment of information. Applying machine learning methods, we discovered that changes in the expressions of anger and joy significantly predict participants' trustworthiness ratings, revealing a broader emotional influence on the perception of information validity among the entire sample.

Conclusions: Our study highlights the complex interplay between emotional responses and political beliefs in shaping the perceived trustworthiness of information, revealing key factors that contribute to the challenge of navigating misinformation. These findings emphasize the need for further research and innovative strategies that account for ideological and emotional influences when tackling misinformation in our increasingly polarized information landscape.

P3-I-199 - Inferring affect from neural activity elicited by monetary incentives

Ryan Yan [1], **Cynthia Wu** [1], **Tara Srirangarajan** [1], **Brian Knutson** [1]

[1] Stanford University

Details

Monetary incentives and outcomes reliably alter neural activity in circuits associated with incentive processing, which has been linked to retrospective affect reports. Researchers have not yet determined whether, however, neural activity can predict affect ratings at the same timescale, on the order of seconds. Healthy subjects (N = 30) completed a modified Monetary Incentive Delay task (Knutson et al., 2001) while undergoing FMRI scanning optimized for subcortical imaging and concurrent pupillometry. During each trial, subjects first saw a monetary cue indicating potential gain or loss, fixated for a delay, motorically responded to a target to gain or avoid loss, and were notified of the outcome. On each trial, subjects were probed about their current affect either during anticipation or outcome. In two trial blocks, subjects rated valence and arousal on a five-point scale, while on the other two blocks, subjects chose one of six listed emotions.

Replicating previous findings (Yan et al., SNE 2023 poster), pupillary dilation during monetary incentive anticipation and outcome predicted self-reported arousal seconds later (β = 0.2). This association was temporally specific to during the anticipatory and outcome phases of the task, but not self-report phases. In ROI analyses, we tested whether neural activity in the Nucleus Accumbens (NAcc) and the

Anterior Insula (AIns) predict affect (Knutson, Katovich, and Suri, 2014). NAcc and AIns activity during anticipation but not outcomes predicted subsequent self-reported arousal but not valence.

For whole brain analyses, positive and negative arousal, calculated from the arousal and valence ratings, were used as parametric regressors to model anticipation; and valence ratings were used a parametric regressor to model outcome. During anticipation, activity in the NAcc, AIns, the dorsal Anterior Cingulate Cortex (dACC), and the ventrolateral Prefrontal Cortex was associated with subsequently reported positive arousal ($p<.001$, uncorrected); whereas activity in the dorsal caudate, the dACC, and AIns was associated with negative arousal. For the outcome phase, self-reported valence was positively associated dorsomedial prefrontal and orbitofrontal cortices activity, but negatively associated with activity in the periaqueductal grey and locus coeruleus. As with pupillary dilation, neural activity during the rating phase was not correlated with self-reported affect. Together, these findings imply that incentive processing synchronizes neural activity and subsequent self-reported affect, informing the construction of tools for inferring affective responses to more dynamic and naturalistic stimuli from neural activity.

P3-I-200 - Bottom-Up and Top-Down Attention in Value-Based Choice

Blair Shevlin [1], **Daniil Luzyanin** [2], **Clayton Hickey** [2], **Arkady Konovalov** [2]

[1] Icahn School of Medicine at Mount Sinai, [2] University of Birmingham

Details

Objective. Value-based choices can be affected both by bottom-up attention (perceptual properties of the stimuli) and top-down attention (goals of the decision maker, such as values of the choice options). Established cognitive computational models of attention suggest that visual attention can drive choices toward salient stimuli if their salience is congruent with the task goals (i.e., choosing options with higher value). Using a simple two-alternative forced-choice paradigm and eye tracking, we decompose the effects of bottom-up and top-down attention by manipulating both the physical properties of the stimuli (adjusting contrast and orientation) and their value to the decision maker (by adjusting rewards assigned to the stimuli).

Methods. On each task trial, subjects (N = 80) chose between two Gabor patches. The patches had four possible orientations tilted from left to right and four possible contrasts. All possible pairs of patches with these 4x4 properties formed 256 trials of a single block. Each of 4 blocks of trials had one of the following tasks: (1) “Choose the brighter patch”; (2) “Choose the darker patch”; (3) “Choose the patch more tilted to the left”; (4) “Choose the patch more tilted to the right.” Additionally, subjects were assigned to one of the two between-subject reward conditions. In the “Fixed reward” condition, they received 1 point for a correct answer and 0 for an incorrect one. In the “Variable reward” condition, they received 1-4 points based on the level of the relevant feature. We recorded subjects’ eye movements at 1000 Hz using an Eyelink 1000 eye tracker.

Results. As preregistered, we found that subjects were more accurate ($p < 0.001$) and faster ($p < 0.001$) when they had to detect the brighter patch. There was a small overall difference (3 p.p.) in accuracy between the binary and scaled reward conditions, but it was not statistically significant ($p = 0.35$). We also

did not find a significant effect of the reward condition on RT ($p = 0.75$). However, there was a significant interaction between the reward condition and task condition ($p = 0.002$), with a strong difference in RT between "choose brighter" and "choose darker" conditions in the binary reward condition and no difference in the scaled reward condition.

Conclusions. Our results demonstrate that bottom-up and top-down attentional systems can potentially have an interactive effect on value-based choice. We plan to develop these findings further by including the analysis of eye-tracking data and applying computational modeling that can simultaneously capture choices, RT, and gaze durations.

P3-I-201 - Attention Allocation under Scarcity: Theory & Experiment

Patrick Sewell [1]

[1] Universitat Pompeu Fabra

Details

This paper provides a theoretical framework and a novel experimental design that allows us to rigorously test the effects of the scarcity mindset on decision making and productivity. Popularized by Mullainathan and Shafir (2013), scarcity is a psychological state where individuals perceive their resources as insufficient relative to their needs. The authors posit that scarcity consumes cognitive resources and impairs decision-making. Despite several field and lab experiments revealing the psychological consequences of scarcity, they do not always study the same outcomes and there is conflicting empirical evidence. Motivated by the far-reaching implications of this theory, I develop a framework that unifies the behavioral findings under one lens and propose a new approach to induce perceived scarcity in the lab.

Since the seminal paper by Shah et al. (2012), the literature has identified three behavioral effects of scarcity: bandwidth tax, trade-off thinking, and tunneling. In this paper, I argue that all these effects can be modeled and explained by an attention allocation problem. I propose a framework where the DM has to decide how to allocate her attention across tasks depending on their expected payoffs and complexity; then show how perceived scarcity in a particular resource can affect this allocation and her choices. Scarcity is introduced via a kinked utility function, formalizing the idea that her marginal utility is much greater if she has less than she feels she needs. Through the model, I characterize how and when the scarcity mindset occurs; furthermore, the model predicts that 1) DM will focus more on tasks where scarcity is perceived, at the expense of performance in other tasks; 2) DM with identical resources may allocate attention differently based on their perceived need threshold; 3) Increased task complexity can have attentional spillovers to other tasks.

I develop a new method of inducing perceived scarcity in the lab to rigorously test the theoretical predictions. Subjects engage in a dual-task scenario: a Knapsack Problem (KP) representing consumer choices and a slider task reflecting productivity. Scarcity is induced by adapting subjects to either high or low budgets in the KP then giving them the same budget, allowing for comparisons when both groups receive identical resources but perceive them differently.

Pilot results, using a similar dual-task scenario with a stock market prediction task, support most theoretical predictions. Despite a small sample size (n=22), the findings align with the model, showing significant effects on attention allocation and decision-making. These preliminary results suggest that the experimental design is robust for studying the nuanced impacts of perceived scarcity.

P3-I-202 - Testing predictions of a model for flexible goal-directed decision-making

Romy Froemer [1], **Chih-Chung Ting** [2], **Sebastian Gluth** [2], **Amitai Shenhav** [3]

[1] University of Birmingham, [2] University of Hamburg, [3] Brown University

Details

Objective. Every day, humans flexibly make a broad range of decisions, including choosing the item they like most or least, or assigning a value to their option set as a whole. We recently showed that a single sequential sampling model could flexibly accommodate these and other types of decisions. We developed a theoretical framework that formalizes the necessary representations that align sequential sampling and evidence accumulation with one's current choice goals. We implemented this framework within an extended leaky competing accumulator model and showed that the model can parsimoniously explain behavior across a range of different choice goals, while also generating predictions for previously untested choice goals. Here we test behavioral predictions of our model.

Methods. To test these predictions, we had 44 participants (37 female, Age: M = 24, SD= 4) choose or appraise option sets under different transformation goals (choose best, choose worst, appraise liking, appraise disliking). Option sets varied in value difference and overall value based on participants prior ratings of the options individually. All option sets were seen once in a choice and once in an appraisal condition. The order of conditions was counterbalanced. Eye movements were recorded during the decision-making and appraisal tasks.

Results. Human behavior matched the predictions of our model. Response times decreased with increasing overall value in choose best and increased in choose worst. During appraisal, response times showed an inverse U-shaped relationship with overall values, with faster response times for more extreme values. Notably, as predicted by our model, this relationship was the same across liking and disliking appraisals. In addition to these predicted behavioral effects, we observed corresponding goal-specific effects on gaze: participants dwelled more on high value options during choose best and more on low value options during choose worst, whereas they showed no value biases during appraisal, consistent with an even gaze allocation across both appraisal conditions.

Conclusion. Our results confirm predictions of our flexible decision-making model. They also shed new light on the flexible value processing mechanisms that enable adaptive decision-making across a wide

variety of goals. Our findings form the basis for dissecting the neural mechanisms underpinning the goal-dependent transformations implemented in our model and offer new targets for understanding when and how decision-making may go awry.

P3-I-203 - The role of attention in choice architecture for promoting sustainable behavior

Barbara Oberbauer [1], **Ulf Hahnel** [2], **Sebastian Gluth** [3]

[1] University Hamburg, [2] Universität Basel, [3] University of Hamburg

Details

To prevent severe consequences of climate change and maintain a habitable planet earth, it is crucial to minimize global warming (IPCC, 2023). To promote sustainable decisions in individuals, choice architecture has been suggested as a powerful tool as it facilitates desired behaviors by creating environments that account for the decision-maker's bounded rationality (Thaler & Sunstein, 2008). However, the effectiveness of such interventions varies and the cognitive mechanisms driving the effects of successful interventions remain poorly understood (Mertens et al., 2022). To unravel these mechanisms, we use a sequential sampling model to reevaluate a large, existing dataset (*N*=792) featuring a multi-attribute consumer choice task.

In the respective study, Mertens et al. (2020) explored the effects of attribute translations, an intervention that highlights different but correlated aspects of an attribute. The authors collected choice and eye-tracking data on product choices as well as information acquisition in the absence and presence of translated attributes. Specifically, participants were asked to choose between washing machines based on information about price, energy consumption, and popularity. In a subsequent session, the energy efficiency was translated into one additional energy attribute. The results affirmed that attribute translations significantly increased ecological product choices, mediated by the amount of dwell time on attributes.

To disentangle the role of attention, we use the multi-attribute attentional DDM from Yang and Krajbich (2023). This model assumes that attention influences choices by discounting the weights of non-attended attributes. Our parameter estimates support the proposed mechanisms of attribute translations, showing a decreased weight for price and an increased weight for energy efficiency in the translation compared to baseline session. Notably, the modeling results also indicate a reduction in the discounting of non-attended attributes in the translation session, suggesting that energy efficiency does not only become more important, but also that participants consider non-attended attributes more.

In conclusion, by using a dynamic choice model that incorporates attention, we were able to provide mechanistic explanations for the effects of attribute translations on sustainability choice.

P3-I-204 - Determinants of context-dependent valuation in artificial and biological neural networks

Kenway Louie [1], **Shreya Sinha** [1], **Dino Levy** [2]

[1] New York University, [2] Tel Aviv University

Details

Objective: In contrast to the assumptions of normative choice theories, empirical decision-making is context-dependent, varying with factors like the composition of the choice set. This context-dependence implies a relative rather than absolute value code, consistent with relative reward representations seen in neural activity. However, it is unknown why context-dependent valuation arises and what function it serves; furthermore, the extent and distribution of relative reward coding across different brain regions is unclear. Here, we hypothesize that relative value coding: (1) arises intrinsically in neural networks trained to make value-guided choices, (2) reflects internal processing constraints, and (3) is governed by external properties of the reward environment.

Methods: We trained simple deep neural networks with gradient descent to make value-guided choices; in primary simulations, value inputs were drawn from multivariate normal distributions with added environmental normal noise. Simulations were implemented in Python and analyzed in Python and Matlab. Following training, relative value coding in hidden layer units was quantified by multiple regression and the calculation of a *relative value index (RVI)*. Across simulation runs, we examined how RVI depended on both internal network (layer size, layer depth, neural nonlinearity, training loss function) and external environmental (reward covariance, input noise, reward distribution) factors.

Results: We find that relative value coding arises naturally in deep networks trained to perform value-guided decision making, evident as a positive RVI in hidden layer units following training. RVI exhibited an anatomical gradient: in multilayer networks, the overall strength of value coding was stronger in later layers but the degree of relativity was highest in early layers, suggesting an initial contextualization of value information prior to subsequent processing. Furthermore, RVI was driven by external correlation, with stronger relativity when reward inputs were more correlated, supporting the hypothesized role of relative coding in reducing redundancy introduced by statistical structure in the world.

Conclusions: These findings suggest that context-dependent valuation serves to implement efficient coding, optimizing performance under biological capacity constraints. Relative value coding arises naturally in artificial network optimization, governed by both internal and external factors. Ongoing work will use the layer-dependent strength of relative valuation as a framework to examine human neuroimaging data, identifying core brain regions in the hierarchical evaluation and decision-making process.

P3-K-205 - Optimal Decision Time

Sota Ichiba [1], **Alex Imas** [2], **Collin Raymond** [3], **Mike Kuhn** [4]

[1] Tilburg University, [2] University of Chicago, [3] Cornell University, [4] University of Oregon

Details

There is a long-standing literature focusing on the role of deliberation in decision making, where models like the evidence accumulation model are used to understand the phenomenon. We critically evaluate the assumption that people correctly anticipate the return to deliberation and optimally pick the response time (RT). To this end, we develop an experiment (N=99) where the participants solve different configurations of the knapsack task. We introduce changes in return to deliberation through two channels -- incentive attached to solving the problem, and complexity of the problem. Our findings suggest polarizing responses from different sources of change in return to deliberation. Increases in return via incentive levels lead to a longer RT. At the same time, increases in return via question difficulty lead to a shorter RT (both $p<0.01$). Simultaneously showing the two-fold pattern robustly demonstrates a violation of rational inattention assumption, as the class of models demand that both of the shocks should affect the RT in the same direction. We argue that this stems from salience of different parts of the return -- people easily see the difference in incentive, but they do not flexibly adjust the RT allocation during the task by learning the difficulty.

P3-L-206 - Improve Theorizing and Value of Empirical Work with an Interactive Tool for Data Simulation

Finn Luebber [1], **Sören Krach** [1]

[1] University of Lübeck

Details

One of the goals of research on decision-making is to understand the complex interplay of various factors contributing to individuals navigating their life and trying to maximize their outcomes. Especially in non-experimental settings, study designs and analysis strategies crucially depend on theoretical reasoning, especially causal assumptions, as different views about how variables interact with each other can require different variables to collect and to statistically adjust for in order to minimize risk of bias and improve precision in the estimate of interest. For a given model, the *degree* of bias and precision hinges upon the strength of variable interaction and their variability (including measurement precision). Thus, estimation of the variability in the estimate of interest (precision) and its expected value (bias) needs a parametric theoretical model.

If data is simulated from this model multiple times, the distribution of results can indicate expected bias, precision, and thus ultimately also power of a planned empirical study. However, existing software packages require quite advanced programming skills and are thus not accessible to all researchers. Thus, we developed a ShinyApp which enables users to construct models as a directed acyclic graph (DAG) in an intuitive interface from which data can be directly simulated.

Within the model, users can add or remove variables, set their error magnitudes and causal effect sizes between variable pairs. The simulated data from that model can subsequently be evaluated within the app: Supported by displayed adjustment sets (variables to adjust for to identify the effect of interest), users can analyze the simulated data using linear or logistic regression and compare the simulated result distribution with the theoretically postulated causal effect size.

Thus, firstly, this tool will direct users to formalize verbal theories, which helps communicating and criticizing theory behind empirical investigations and strengthening the theoretical-empirical cycle. Secondly, the implications of this formalization are directly fed back, transmitting important information prior to actual empirical work, for instance: Is the effect of interest identifiable, do additional variables have to be measured, which sample size is needed, which measures need improvement? For these reasons, this app facilitates theoretical work in the theoretical-empirical cycle and also has the potential to increase the value of empirical work in many areas of social and life science, including human decision-making and neuroeconomics.