# The 20th International Conference on Social Dilemmas

**Leiden University**

July 2 – 5, 2024

https://socialdilemma.com/icsd2024

icsd2024@fsw.leidenuniv.nl

Emergency phone: +31 (0)71 527 6705

**Local Organizing Committee**

Erik de Kwaadsteniet (Chair), Carsten de Dreu, Eric van Dijk, Leticia Rettori Micheli, Dorothee Mischkowski, Niels van Doesum, Wolfgang Steinel, Angelo Romano, Jaime Vigil Escalera Sanchez, Maria Lojowska, Welmer E. Molenmaker (Program chair), Arjaan Wit, Milena Poulina, Conny Binnendijk

socialdilemma.com

Universiteit Leiden
The Netherlands

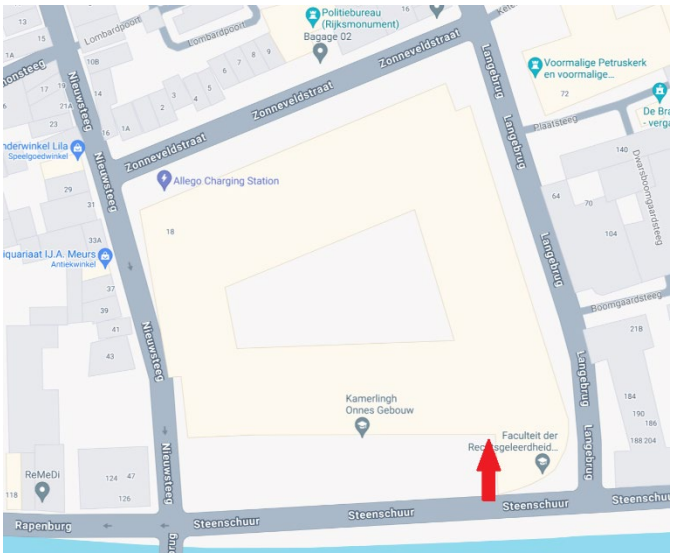
| | Monday 1 July | | Tuesday 2 July | | Wednesday 3 July | | Thursday 4 July | | Friday 5 July | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Room 1 | Room 2 | Room 1 | Room 2 | Room 1 | Room 2 | Room 1 | Room 2 | Room 1 | Room 2 |
| 08:00 | | | Registration and walk-in | | | | | | | |
| 08:15 | | | | | | | | | | |
| 08:30 | | | | | Walk-in | | Walk-in | | Walk-in | |
| 08:45 | | | | | | | | | | |
| 09:00 | | | Opening | | Session 5 | Session 6 | Session 11 | Session 12 | Session 15 | Session 16 |
| 09:15 | | | | | | | | | | |
| 09:30 | | | Session short 1 | | | | | | | |
| 09:45 | | | | | | | | | | |
| 10:00 | | | | | | | | | | |
| 10:15 | | | | | | | | | | |
| 10:30 | | | 30 min. break | | 30 min. break | | 30 min. break | | 30 min. break | |
| 10:45 | | | | | | | | | | |
| 11:00 | | | Session 1 | Session 2 | Session 7 | Session 8 | Session 13 | Session 14 | Session 17 | Session 18 |
| 11:15 | | | | | | | | | | |
| 11:30 | | | | | | | | | | |
| 11:45 | | | | | | | | | | |
| 12:00 | | | | | | | | | | |
| 12:15 | | | | | | | | | | |
| 12:30 | | | 90 min. break | | 90 min. break | | 90 min. break | | Closing (and awards) | |
| 12:45 | | | | | | | | | | |
| 13:00 | | | | | | | | | | |
| 13:15 | | | | | | | | | | |
| 13:30 | Preconference | | | | | | | | | |
| 13:45 | | | | | | | | | | |
| 14:00 | | | Session 3 | Session 4 | Session 9 | Session 10 | Keynote: **Mariska Kret** | | | |
| 14:15 | | | | | | | | | | |
| 14:30 | | | | | | | | | | |
| 14:45 | | | | | | | | | | |
| 15:00 | | | | | | | announcements | | | |
| 15:15 | | | | | | | Social events (by registration only) | | | |
| 15:30 | | | 30 min. break | | 15 min. break | | | | | |
| 15:45 | | | | | Session short 2 | | | | | |
| 16:00 | | | Keynote: **Jörg Gross** | | | | | | | |
| 16:15 | | | | | | | | | | |
| 16:30 | | | | | 15 min. break | | | | | |
| 16:45 | | | | | | | | | | |
| 17:00 | Registration and welcome reception | | Poster session (with drinks) | | Keynote: **Alan Sanfey** | | | | | |
| 17:15 | | | | | | | | | | |
| 17:30 | | | | | | | | | | |
| 17:45 | | | | | | | | | | |
| 18:00 | | | | | | | Transport to conference dinner | | | |
| 18:15 | | | | | | | | | | |
| 18:30 | | | | | | | | | | |
| 18:45 | | | | | | | | | | |
| 19:00 | | | | | | | Conference dinner | | | |
| ... | | | | | | | | | | |
| ... | | | | | | | | | | |
| 23:00 | | | | | | | | | | |

## Conference location

The conference takes place at the **Kamerlingh Onnes Gebouw** at Leiden University. This beautiful building from the 19th century was originally the physics laboratory of Leiden University (famous scientists like Marie Curie and Albert Einstein have done research here in the past), and is currently the Law Faculty building.

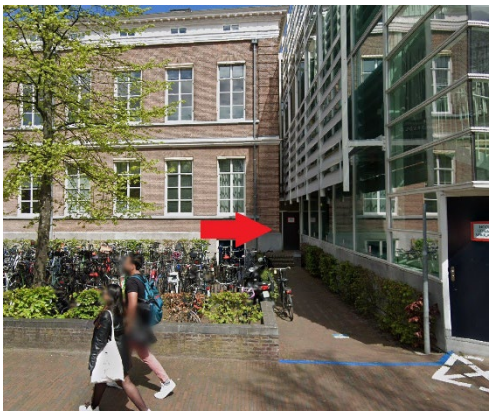
The building is located in the historical center of Leiden, only a 20 minutes' walk from Leiden Central Station.

Address: Steenschuur 25, 2311 ES Leiden




Main entrance

The main entrance of the Kamerlingh Onnes Gebouw cannot be used due to maintenance work. To enter the building, you must use the temporary entrance on the same side as the main entrance, in the right corner when facing the building (see images below).



## Wi-Fi

You can connect to the eduroam Wi-Fi network at the conference location in three ways:

- If your university or institution has an eduroam Wi-Fi network too, you should have automatic Wi-Fi access with your university account.
- You can request temporary eduroam Visitors Access via SMS. See the Program Booklet you received via email with instructions, or the instructions at the registration desk.
- You can get a personal Wi-Fi username and password for temporary eduroam Visitors Access at the registration desk.

| Regular talks | |
|---|---|
| Tuesday July 2, 2024 | |
| **Session 1: Personality** | |
| 11:00 – 11:15 | **Natalie Popov** – The core tendencies underlying prosocial behavior: Testing a person situation framework |
| 11:15 – 11:30 | **Alicia Seidl** - Who turns a blind eye? – Investigating the dispositional basis of unethical loyalty |
| 11:30 – 11:45 | **Simona Cicognani** - Cooperating when the end is near: the impact of cognitive ability and task complexity |
| 11:45 – 12:00 | **Isabel Thielmann** - The dispositional basis of human prosociality: One common core or multiple related factors? |
| 12:00 – 12:15 | **Busra Yelbuz** - Who helps whom? An investigation of individual differences in selective prosociality |
| 12:15 – 12:30 | **Ranran Li** - Revisiting Situational Strength: Do Strong Situations Restrict Variance in (Cooperative) Behaviors? |
| **Session 2: Evolution of cooperation** | |
| 11:00 – 11:15 | **Ilan Fischer -** The influence of similarity perceptions on the evolution of cooperation and confrontation |
| 11:15 – 11:30 | **Christian Hilbe -** Efficiency and resilience of cooperation in asymmetric social dilemmas |
| 11:30 – 11:45 | **Anil Yaman -** A computational model of social sanctioning for encouraging division of labor |
| 11:45 – 12:00 | **Ro'i Zultan -** The evolution of group reciprocity |
| 12:00 – 12:15 | **Mohammad Salahshour -** The evolution of risk-aversion and human risk-taking behavior in interacting public goods |
| 12:15 - 12:30 | **Marco Colnaghi -** Navigating social dilemmas: adaptations to infer fitness interdependence promote the evolution of cooperation |
| **Session 3: Climate change & Health threats** | |
| 14:00 – 14:15 | **Lidor Krava-Farkovitsh -** Who is going to save the planet? The relationship between Passive risk-taking, Social attitudes, and Pro-environmental behaviors |
| 14:15 – 14:30 | **Simon Columbus -** The Social Dilemma of Climate Policy |
| 14:30 – 14:45 | **Yael Mintz -** Climate Change Ignorance: Passive Risk Taking, Self-Efficacy, and Mitigation Intentions |
| 14:45 – 15:00 | **Sakshi Kasa Prasad -** Can Climate Clubs Curb Climate Change? - An Experimental Study |
| 15:00 – 15:15 | **Manfred Milinski -** Climate extreme events are enforced by extortionate free-riders |
| 15:15 – 15:30 | **Mitchell Matthijssen -** Getting vaccinated for pro-social reasons: when does it work? |
| **Session 4: Intergroup conflict** | |
| 14:00 – 14:15 | **Qinyu Xiao -** Individual exploitation in intergroup conflicts |
| 14:15 – 14:30 | **Dora Simunovic -** To Share or to Freeride? Public Goods in Heterogeneous Societies |
| 14:30 – 14:45 | **Nobuyuki Takahashi -** Are cooperators more likely to attack out-group members than defectors? – The second experiment |

| | |
|---|---|
| 14:45 – 15:00 | **Tamar Kugler -** Do the Powerful Compete or Cooperate? An Intergroup Exploration of Power |
| 15:00 – 15:15 | **Lennart Reddmann -** Unpredictable Futures, Parochial Pro-Sociality, and Intergroup Conflict |
| 15:15 – 15:30 | **Luuk Snijder -** Leader Rhetoric Escalates Intergroup Conflict |
| Wednesday July 3, 2024 | |
| **Session 5: Sustainability** | |
| 09:00 – 09:15 | **Giulia Priolo** - Who, how much and when: effect of endogenously suggested rules and inequality on a common resource game |
| 09:15 – 09:30 | **Marco Janssen** - What makes communication effective to govern shared resources? |
| 09:30 – 09:45 | **Eliran Halali** - Binding the Future: Far-looking altruism boosts long-term sustainability |
| 09:45 – 10:00 | **Jan K. Woike** - Over-harvesting like there is no tomorrow: Evidence from dilemma-free common pool games |
| 10:00 – 10:15 | **Jan Hausfeld** - The Effect of Ecolabels and Attention on Producer's Sustainable Decision-Making |
| 10:15 – 10:30 | **Erik de Kwaadsteniet** - Can exclusion prevent the Tragedy of the Commons? An experiment comparing excludable vs. non-excludable resources |
| **Session 6: Trust** | |
| 09:00 – 09:15 | **Zeyu (Arthur) Xue** - Trust in Gender and Profession—A Social-Category Account |
| 09:15 – 09:30 | **Jan B. Engelmann** - Emotional determinants of trust behavior |
| 09:30 – 09:45 | **Mathias Twardawski -** The Impact of Reputation on Trust: Examining Individual Differences in Using Third-Party Evaluations in Cooperation Contexts |
| 09:45 – 10:00 | **Stephan Nuding -** They want and they can: Do decisions in shared social dilemmas depend not only on expectations of others' goodwill but also on others' competence? |
| 10:00 – 10:15 | **Wojtek Przepiorka -** The prevalence and magnitude of the correlation between generosity and trustworthiness: a meta-analysis of economic game experiments |
| 10:15 – 10:30 | **Eyal Ert -** Do people have a "trust propensity"? On the relation between common elicitation methods of trust |
| **Session 7: Cross-Cultural** | |
| 11:00 – 11:15 | **Giuliana Spadaro -** Institutional trust and closeness-based favoritism across 25 societies |
| 11:15 – 11:30 | **Astrid Hopfensitz -** The impact of loneliness on economic trust: experimental evidence from 27 European countries |
| 11:30 – 11:45 | **Kristen Syme -** Social Control across 60 Societies: self-interested norm enforcement versus strong reciprocity |
| 11:45 – 12:00 | **Ori Weisel -** Corrupt Collaboration in 20 Societies |
| 12:00 – 12:15 | **Catherine Molho -** Guilt- and Shame-Driven Prosociality Across Societies |
| 12:15 – 12:30 | **Vanessa Clemens -** Antecedents and consequences of cross-national social preferences |

| Session 8: Conflict & Competition | |
|---|---|
| 11:00 – 11:15 | **Johann M. Majer -** Unveiling the hidden impact of the conflict structure on parties' settlements. A field-level experimental black-box approach from 1971 to 2021 |
| 11:15 – 11:30 | **Laura E. M. Stalenhoef -** Interpersonal conflict resolution crowds out forward-looking decision making |
| 11:30 – 11:45 | **Annabel Losecaat Vermeer -** Testosterone affects learning of implicit social dominance hierarchies through competitive interactions |
| 11:45 – 12:00 | **Shuxian Jin -** The Cultural Logic of Honor, Competition and Coordination |
| 12:00 – 12:15 | **Uri Zak -** Expected outcomes and risk-taking in competitive contexts: A large-scale analysis of gambits in tournament chess |
| 12:15 – 12:30 | **Ruthie Pliskin -** On the Context-Specificity of Ideological Asymmetries in (Intergroup) Mistrust and Aggression |
| **Session 9: Prosociality in the wild** | |
| 14:00 – 14:15 | **Federica Maria Raiti -** Preferences for Status Quo, Adaptation or Mitigation in Collective Risks Situations |
| 14:15 – 14:30 | **Gianluca Grimalda -** Market integration, egalitarianism, and reward of merit: An experimental analysis from Papua New Guinea small-scale societies |
| 14:30 – 14:45 | **Sylvia. Y. Xu -** Highlighting the Collective Harm: Tackle Illicit Drug Use in the Netherlands with Moral Appeals |
| 14:45 – 15:00 | **Tom Gordon-Hecker -** Charitable Donation Theories in the Wild: Evidence from a Large Online Donation Platform |
| 15:00 – 15:15 | **Thomas Rittmannsberger -** Cooperation and Punishment in the general population: Evidence from a representative experiment in Germany. |
| 15:15 – 15:30 | **Hagai Rabinovitch** - Psychological Mechanisms Underpinning People's Willingness to Vaccinate Against Future Viruses |
| **Session 10: Information** | |
| 14:00 – 14:15 | **Katharina Reher -** Does ignorance love company? Social malleability of information avoidance and decision-making. |
| 14:15 – 14:30 | **Simone Righi -** Impact of Information Disclosure on Corruption and Cooperation: A Public Goods Game Approach |
| 14:30 – 14:45 | **Linh Vu -** Giving (in) to help an identified other |
| 14:45 – 15:00 | **Arkady Konovalov -** Facilitating Cooperation by Manipulating Attention |
| 15:00 – 15:15 | **Lina Koppel -** Comprehension in Economic Games |
| 15:15 – 15:30 | **Shira Garber-Lachish -** Negotiators have the wrong model of their counterparts: What really motivates negotiators' behavior? |
| Thursday July 4, 2024 | |
| **Session 11: Inequality & Fairness** | |
| 09:00 – 09:15 | **Rémi Suchon** - The rich, the poor and strength of Inequality: evidence from a meta-dataset of public good games experiments. |
| 09:15 – 09:30 | **Leticia Micheli** - Moving up? The effect of economic mobility on giving behavior |
| 09:30 – 09:45 | **Jing Lin** - The Impact of Economic Inequality on Charitable Behavior |
| 09:45 – 10:00 | **Leon P. Hilbert** - Pro-sociality of the financially affluent and deprived |

| | |
|---|---|
| 10:00 – 10:15 | **David Munguia Gomez** - People prefer to address inequalities by reducing disadvantage over advantage |
| 10:15 – 10:30 | **Ankush Asri** - Affirmative Action: Within-group Inequality in Competitive Environments |
| **Session 12: Reputation & Communication** | |
| 09:00 – 09:15 | **Terence Daniel Dores Cruz** - The Costs and Benefits of Gossip |
| 09:15 – 09:30 | **Tiffany Matej Hrkalovic** - Partner Perceptions During Brief Online Interactions Shape Partner Selection and Cooperation |
| 09:30 – 09:45 | **Marcus Krellner -** Words are not Wind - How Joint Commitment and Reputation Solve Social Dilemma |
| 09:45 – 10:00 | **Andrea Marietta Leina -** Cooperation in the 'Helping Game': Good Standing or Image Scoring? |
| 10:00 – 10:15 | **Jacobus Martin Smit -** The dynamics of universal cooperation with reputations |
| 10:15 - 10:30 | **Wei Zhang -** Conformity versus credibility: A coupled rumor-belief model |
| **Session 13: Ingroup-Outgroup** | |
| 11:00 – 11:15 | **Natalie Struwe -** The role of strategic uncertainty in collective risk social dilemmas with donors |
| 11:15 – 11:30 | **Filippo Toscano -** Not all groups are created equal: Unequal abilities to cooperate within groups increases inequality and decreases group-transcending cooperation |
| 11:30 – 11:45 | **Hannes Rusch -** Groupiness over time: A longitudinal lab-in-the-field experiment |
| 11:45 – 12:00 | **Hirotaka Imada -** Group-bounded indirect reciprocity and in-group favouritism: recent advancements and future directions |
| 12:00 – 12:15 | **Angelo Romano -** Humans Exhibit both Parochialism and Nastiness within Groups |
| 12:15 - 12:30 | **Laura C. Hoenig -** How Group Cooperation Generates Intergroup Conflict |
| **Session 14: Social norms** | |
| 11:00 – 11:15 | **Annelie Bruning -** Measuring social norms in surveys: The role of question sequence, reference group, and context information in gender norm inquiries. |
| 11:15 – 11:30 | **Jantsje Mol -** Avoidance of Altruistic Triggers: Empathy versus Social Pressure |
| 11:30 – 11:45 | **Giulia Andrighetto -** Risk, sanctions and norm change: the formation and decay of social distancing norms |
| 11:45 – 12:00 | **Dorothee Mischkowski -** The interplay between low- and high-cost cooperation |
| 12:00 – 12:15 | **Zvonimir Bašić -** Personal norms — and not only social norms — shape economic behavior |
| 12:15 - 12:30 | **Shacked Avrashi -** Norm-based and Self-relevant Perceptions of Cooperation |
| Friday July 5, 2024 | |
| **Session 15: Digital interactions** | |
| 09:00 – 09:15 | |

| | |
|---|---|
| 09:15 – 09:30 | **Yehor Hrymchak** - Time and Ties in Moral Social Dilemmas. How Temporal Distance and Personal Closeness Affect Dishonesty and Moral Judgment |
| 09:30 – 09:45 | **Andreas Orland** - Playing Prisoner Dilemma Games with LLMs |
| 09:45 – 10:00 | **Nils Köbis** - Ethical Risks of Algorithmic Delegation |
| 10:00 – 10:15 | **Claudia Keser** - The Economic Effects of Remote-Bargaining vs. In-Person Bargaining |
| 10:15 - 10:30 | **Margarita Leib** - Corrupted by Algorithms? How AI-generated and Human-written Advice Shape (Dis)honesty |
| **Session 16: Reciprocity** | |
| 09:00 – 09:15 | |
| 09:15 – 09:30 | **Pierce Gately** - Rule Following and Cooperation |
| 09:30 – 09:45 | **Sebastian Grüneisen** - Young children protect rule-breakers whom they owe a favor |
| 09:45 – 10:00 | **Ben Grodeck -** Cooperating Across Generations: Experimental Evidence of Reciprocal Cooperation and Intergenerational Exchange |
| 10:00 – 10:15 | **Ryutaro Mori -** Proactive Cooperation and Reciprocation in Real-time PD |
| 10:15 - 10:30 | **Marcel Lumkowsky -** Cooperation and strategy choice in the infinitely repeated prisoner's dilemma when players can cheat |
| **Session 17: Punishment & Reward** | |
| 11:00 – 11:15 | **Julien Lie-Panis -** The repeated punishment game explains why, and when, we seek revenge |
| 11:15 – 11:30 | **Pat Barclay -** Monetary sanctions are more effective than (dis)approval in maintaining long-term cooperation |
| 11:30 – 11:45 | **Ro'i Zultan -** Public feuds and cooperation |
| 11:45 – 12:00 | **David Munguia Gomez -** Luck that builds merit |
| 12:00 – 12:15 | **Jolien van Breen -** Dimensions of Transgressive Social Interaction: A conjoint experiment |
| 12:15 - 12:30 | |
| **Session 18: Games & Game Theory** | |
| 11:00 – 11:15 | **Ivan Romic -** Game theory of the hometown tax system |
| 11:15 – 11:30 | **Yohsuke Murase -** Indirect reciprocity with stochastic and dual reputation updates |
| 11:30 – 11:45 | **Marta C. Couto -** The evolution of boundedly rational learning in games |
| 11:45 – 12:00 | **Nikoleta E. Glynatsi -** Reactive strategies with longer memory |
| 12:00 – 12:15 | **Jorge Peña** - Cooperative dilemmas with binary actions and multiple players |
| 12:15 – 12:30 | **Bryan Bruns -** When games get worse |

| Short talks | |
|---|---|
| Tuesday July 2, 2024 | |
| **Short session 1** | |
| 09:30 – 09:35 | **Nicole Casali** – Can we measure trait morality? Revisiting the Moral Character Questionnaire and introducing a new, integrative self-report scale |
| 09:35 – 09:40 | **Nobuhiro Mifune** – Power imbalance leads to out-group aggression |
| 09:40 – 09:45 | **Dianna Amasino** – Fairness in the face of changing inequality |
| 09:45 – 09:50 | **Daniel Balliet** – Relationship Interdependence Influences Prosocial Behavior Across Ego Networks |
| 09:50 – 10:00 | Questions |
| 10:00 – 10:05 | **Yanyan Chen** – Reputation-based Trust and Cooperation: When and How does Dishonest Reputation Upgrading Backfire? |
| 10:05 – 10:10 | **Matthieu Legeret** – Does Realism Affect Behaviors in Moral Dilemmas? |
| 10:10 – 10:15 | **Isamu Okada** – Reputation dynamics in divided societies: Image updating in indirect reciprocity under private assessment |
| 10:15 – 10:20 | **Eladio Montero Porras** – Prosocial and Self-serving default choices in the common pool resource dilemma are persuasive but not persistent |
| 10:20 – 10:30 | Questions |
| Wednesday July 3, 2024 | |
| **Short session 2** | |
| 15:45 – 15:50 | **Yukari Jessica Tham** – No “cooperation for reputation” in highly asymmetric step-level public goods games |
| 15:50 – 15:55 | **Ryoichi Onoda** – Intergroup vicarious retribution caused by prosocial punishment in social dilemmas |
| 15:55 – 16:00 | **Judit Mokos** – Wanna solve the climate crisis? End inequality – The potential of unequal loss in a climate game kills trust and causes sinking together |
| 16:00 – 16:05 | **Rie Mashima** – What kind of information are people willing to spread? Manipulating information credibility to examine information transfer bias in situations of indirect reciprocity |
| 16:05 – 16:15 | Questions |
| 16:15 – 16:20 | **Bing Jiang** – Are Christians More Forgiving and Less Greedy? Evidence from a Power-to-take Game Experiment |
| 16:20 – 16:25 | **Paul A. M. van Lange** – How do Impressions of Age, Ethnicity, Sex, and Social Class simultaneously Affect Cooperation? |
| 16:25 – 16:30 | **Richardt R. S. F. Hansen** – Managing the Commons: The role of Political Orientation and Framing on Cooperation in a Common-Pool Resource Dilemma |
| 16:30 – 16:35 | **Eugenia Polizzi** – Social Corrections: Nudging norm enforcement against fake news sharing |
| 16:35 – 16:45 | Questions |

| Posters | |
|---|---|
| Tuesday July 2, 2024 | |
| **17:00 – 18:30** | |
| 1. | **Alejandro Hirmas** - Learning the value of Eco-Labels: The role of information in sustainable decisions |
| 2. | **Wei Zhang** - Is cooperation sustained under increased mixing in evolutionary public goods games on networks? |
| 3. | **Toby Handfield** - Cooperative dilemmas in rational debate |
| 4. | **Libera Ylenia Mastromatteo** - Retrospective Self-Reported Adversities and Family Unpredictability Over the Course of Childhood Uniquely Predicts Cooperative Behavior during Adulthood: The Moderating Role of Sensitivity to The Environment |
| 5. | **Ge-yang Chen** - An agent-based model of stochastic punishment by authorized third parties in public goods game: The interplay between different justice concerns and reputation-based migration |
| 6. | **Yasuyuki Kudo** - Machine-learning Approaches for Meta-analytic Estimates of Important Predictors: Analysis of Cooperation in Social Dilemmas |
| 7. | **Caetano Franco** - A framework to explore social norms in co-management of natural resources in a multi-level structure |
| 8. | **Gary Ting Tat Ng** - The moral dilemma and social dilemma of autonomous vehicles: The role of perceived responsibility |
| 9. | **Laura Tietz** - Reciprocal Cheating in Children – Children Break the Rules to Return a Favor |
| 10. | **Xueting Zhang** - The Effects of Emotions on Moral Reactions: A Meta-Analysis |
| 11. | **Hagai Rabinovitch** - When Ignorance is a Curse: Being Blind to Irrelevant Information Compromises Selection Decisions |
| 12. | **Shogo Mizutori** - Does social value orientation moderate the effect of reaction time manipulation on cooperation? |
| 13. | **Tycho van Tartwijk** - The Impact of Between-Group Interaction on Within-Group Cooperation: A Meta-analysis |
| 14. | **Maximilian Schmitt** - Understanding exploitative behavior: Microtheory and evidence |
| 15. | **Shacked Avrashi** - Revealing the Interaction Between Strategic Properties and Intervention Opportunities: A Novel Taxonomy of Two-by-two Games |
| 16. | **Nora Hangel** - Navigating Publication Bias in Judgment and Decision Making (JDM): A Philosophical Analysis of Scientists' Strategies and Perceptions |
| 17. | **Aleksandra Lazic** - Communicating individual benefits promotes vaccination intention in the absence of strong social norms: A preregistered online experiment |
| 18. | **Sakura Ono** - The reputation consequences of punishment – Comparison of give-some and take-some games - |
| 19. | **Julia Teufel** - Mapping perceptions of exploitativeness across situations and interactions |
| 20. | **Jurgis Karpus** - Gender bias in human collaboration with artificial intelligent agents |
| 21. | **Rafael Nunes Teixeira** - The Role of Roles: The Impact of Roles on Behavior and Norms in a Public Good Game |
| 22. | **Irenaeus Wolff** - Whose Norms to Follow? A Field Experiment on Both Sides of a Border |
| 23. | **Kaede Maeda** - Intuitive cooperation across group boundaries in a minimal group paradigm |
| 24. | **Jantsje Mol** - An Experimental Test of Risk Perceptions under a New Hurricane Classification System |

| | |
|---|---|
| 25. | **Hirofumi Hashimoto** - Intuitive cooperation in a one-shot prisoner's dilemma game with gain-loss frames: Revisiting the social exchange heuristic hypothesis |
| 26. | **Amanda M. Lindkvist** - Do monetary incentives matter for measuring social preferences? A comparison of full-payment, random-payment, and hypothetical-choice incentive schemes in five economic games |
| 27. | **Matthieu Légeret** - Prebunking Conspiracy Theories |
| 28. | **Maxime Bourlier** - Kill some to save many? When I feel depressed, I'll think about it. |
| 29. | **Teodora Spiridonova** - Protected, but at what cost? Investigating the potential side-effects of inoculating against misinformation |
| 30. | **Patricia Kanngiesser** - Roll for the future: Introducing a novel climate change game |
| 31. | **Kyra Selina Hagge** - Together for a common cause? Cooperative tendencies in transdisciplinary research groups aimed at solving water quality and quantity issues in Eastern North Carolina |
| 32. | **Jaime Vigil Escalera Sanchez** - The development of social prediction during adolescence |
| 33. | **Adam W. Stivers** - Correlates of Social Preferences |
| 34. | **Olivia Seubert** - The evaluation of third-party punishment depending on its type, severity, and interpersonal hierarchy |
| 35. | **Nicolas Coucke** - The dual influence of interpersonal bonds on the cognitive processes of disobedience |
| 36. | **Fredrik Jansson** - Trust-based information filtering can form polarising group identities |
| 37. | **Fredrik Jansson** - Aversive medical treatments signal a need for support |
| 38. | **Yoko Kitakaji** - Cooperation in Nested Social Dilemmas: The Role of Pooled Punishment |

# Social Events and Conference Dinner

When registering for the conference, you could buy a ticket for an optional excursion of your choice. The conference dinner is included in the conference fee.

**Optional excursions**

- **Brewery tour and beer tasting** – We meet at the conference desk at 15:15 and walk approx. 1km to the brewery (Brouwerij Pronck, Langegracht 90D, Leiden). After the tasting, a bus will pick us up at the brewery and take us to the Conference Dinner. See also the map on the next page.
- **Boat tour** – We meet in Room 1 at 15:15 and start with a short introduction by Arjaan Wit ("Dry feet, wet feet; consequences of climate change in this region of the Netherlands"). Next, we walk to the boats (in front of the Kamerlingh Onnes Gebouw). After the boat tour, we leave the boats near the Parking Haagweg (Haagweg 8, Leiden) where busses will wait to take us to the Conference Dinner. See also the map on the next page.

**Conference Dinner**

Our conference dinner will be a beach party including a barbecue and a DJ. This takes place at the best beach bar of South Holland, Surf en Beach at Katwijk aan Zee (starting at 19:00; Address: Boulevard Zeezijde 9, Katwijk aan Zee).

**How to get to the Conference Dinner?**

- If you booked an optional excursion, a bus will take you directly to the beach bar.
- If you have not booked any of the optional excursions, you can make use of the organized bus transportation or public transport (see below).

Organized bus transportation

We have organized bus transportation to take you directly to the beach bar. The busses leave from Parking Haagweg (Haagweg 8, Leiden) at 18:00. So, make sure you arrive on time.

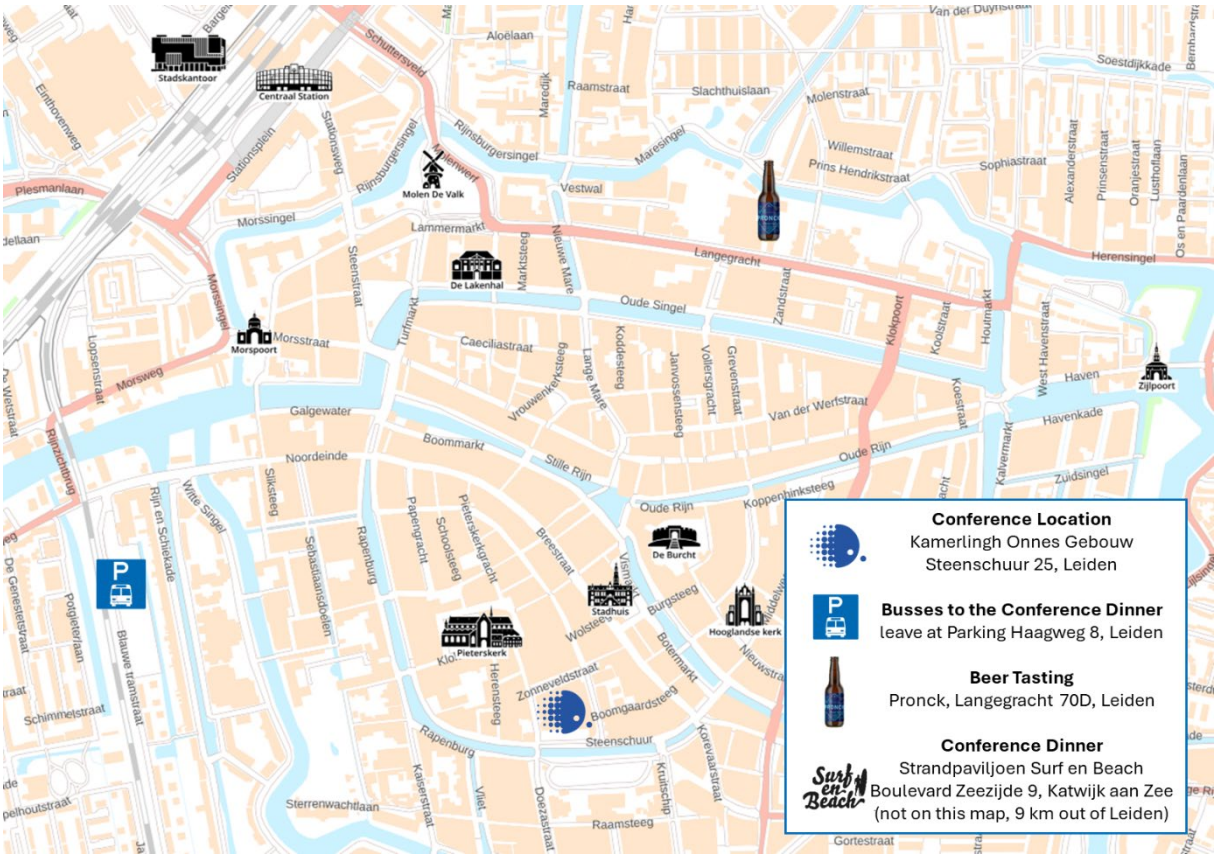
Public transport

You can travel to the beach yourself (earlier) by public transport. Busses to Katwijk aan Zee leave approx. every 20 minutes from Bus Station Leiden Central, Platform D. Take line Arriva R-NET Bus 431 direction Katwijk Boulevard Zuid and get off at Katwijk Centrum. It's an 18-minute ride plus 8 minutes' walk to the beach bar (see the map on the next page). R-NET Bus 431 busses back to Leiden leave approx. every 30 minutes until 0:30 from Bus Stop Katwijk Centrum. Plan your journey on https://9292.nl/en.

**How to get from the Conference Dinner back to Leiden?**

Our busses pick us up on the boulevard in front of the beach bar. One bus will leave at 22:00, and three busses will leave at 23:00. Alternatively, you can take public transport back (see above and the map on the next page).

## Map of Leiden




## Map of Katwijk aan Zee

Note that the organized bus transportation leaves from the boulevard in front of the beach bar.

Keynote | Room 1

Tuesday July 2 16:00 – 17:00

## From self-reliance to intergroup cooperation: Laboratory studies on the dynamics of social organization

**Jörg Gross** (University of Zürich)

Humans are considered incredibly cooperative, even though cooperation is always threatened by temptations to free-ride. To maintain cooperation, social dilemmas emerge on different levels of social organization. In this talk, I will present a series of laboratory studies that aim to elucidate various social dynamics that help or hinder individuals in transitioning from (i) individual-level problem-solving to group cooperation, (ii) group cooperation to intergroup cooperation, and (iii) small-scale to large-scale cooperation. Results highlight how the ability to shape the 'rules of the game' or the choice of whom to interact with can enable groups to not only stabilize group, but also intergroup cooperation through mechanisms like social interdependence and indirect reciprocity.

Keynote | Room 1

Wednesday July 3 17:00 – 18:00

## Multiple motivations in moral choice: computational and neural approaches

**Alan Sanfey** (Radboud University)

Our lives consist of a constant stream of decisions, from the mundane to the highly consequential. The standard approach to studying decision-making has been to examine choices with clearly defined probabilities and outcomes, however it is challenging to extend these decision models to situations where one's outcomes depend on the choices of others, and their outcomes on you. This class of 'social' decision-making better approximates many of our real-life choices, and these social interactive scenarios reveal important motivations, other than simple economic gain, that guide our decisions in a systematic fashion. For example, people consistently value prosocial acts more than standard economic models predict, though importantly, different people have different reasons for acting altruistically. In this talk I will outline an experimental approach using functional brain imaging and computational modelling to observe how players decide in real, consequential, social contexts, and how we can assess what motivations underlie these choices.

Keynote | Room 1

Thursday July 4 14:00 – 15:00

## Emotion Processing in Homo and Pan

**Mariska Kret** (University of Leiden)

Throughout evolution group-living species, including both humans and non-human primates, have developed the ability to quickly recognize and respond to the emotional expressions of others. When individuals unconsciously mimic the emotional expressions of interaction partners, they experience a reflection of those emotions, which guides their social behaviors such as trust, cooperation, and even help form romantic relationships. While much of the research on emotions has focused on facial expressions in humans, there is still a lot to learn about body language and subtle physiological responses like pupil dilation and blushing. In this talk, I take a comparative approach, examining similarities and differences in how humans and great apes express and perceive emotions. Ultimately, our goal is to deepen our understanding of emotional expressions and how we interpret each other's emotions, which can contribute to stronger social connections and relationships.

Session 1: Personality | Room 1

Tuesday July 2, 2024 11:00 - 11:15

**The Core Tendencies Underlying Prosocial Behavior: Testing a Person-Situation Framework**

Natalie Popov | Isabel Thielmann

Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany | Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany

Rationale: Personality is a consistent predictor of prosocial behavior. According to a recently proposed theoretical framework, different traits should explain prosocial behavior in different situations. Specifically, four key affordances can be distinguished in interdependent situations: a possibility for exploitation, a possibility for reciprocity, temporal conflict between short-term and long-term interests, and dependence on others under uncertainty. In turn, each of these affordances should activate a certain personality trait class and thereby allow the expression of corresponding traits in prosocial versus selfish behavior. While this theoretical framework was tested meta-analytically, we experimentally tested its key proposition that each of four "core tendencies" (i.e., the shared variance of traits from the same trait class) predicts prosocial behavior in the presence of a specific situational affordance.

Methods: We used a large and demographically diverse dataset (N = 2,479) including measures of various personality traits that were specifically selected to broadly represent each of the four core tendencies as well as six incentivized economic games assessing prosocial behavior in different social situations. Correspondingly, the games in our dataset primarily provided one of the four key situational affordances and thus involved specific aspects of prosocial behavior. Using bifactor modelling, we extracted four latent core tendencies and tested their predictive validity for prosocial behavior.

Results: We found mixed support for the theoretically-derived, pre-registered hypotheses. Whereas the core tendencies of unconditional concern for others' welfare and beliefs about others' prosociality specifically predicted prosocial behavior in situations involving a possibility for exploitation and dependence under uncertainty, respectively, evidence was relatively weak for conditional concern for others' welfare and self-regulation predicting prosocial behavior in the presence of a possibility for reciprocity and temporal conflict, respectively. However, all bifactor models yielded a good fit to the data and indicated that scales of the same trait class share a meaningful amount of variance, suggesting that there is an underlying common core representing the four suggested core tendencies.

Conclusion: Different features of social situations may activate different personality traits to influence prosocial behavior, but more research is needed to fully understand these person-situation-interactions. This is particularly evident for the core tendency of self-regulation, and less so for conditional concern for others welfare where further tests of the affordance-based framework of individual differences in prosocial behavior are necessary to provide a deeper understanding of the personality traits that drive prosocial behavior in different situations.

Session 1: Personality | Room 1

Tuesday July 2, 2024 11:15 - 11:30

**Who turns a blind eye? – Investigating the dispositional basis of unethical loyalty**

Alicia Seidl | Benjamin E. Hilbig | Isabel Thielmann

Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany | RPTU Kaiserslautern-Landau, Landau, Germany | Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany

Often enough, unethical acts are witnessed by bystanders who can either blow the whistle about the observed transgression or cover it up. As an extreme example with detrimental consequences for individuals and society at large, consider individuals who cover up influential politicians' unethical behavior for hush money. Such acts of (lying for the purpose of) covering up transgressors' unethical behavior have been referred to as unethical loyalty: They are unethical because they protect a transgressor from potential sanctions while at the same time being loyal or prosocial towards the transgressor. Interestingly, from a personality perspective, unethical and prosocial behavior are usually predicted by opposing levels of the same traits, meaning that they do not tend to go hand in hand. Most prominently, the HEXACO personality trait Honesty-Humility has been consistently negatively linked to unethical behavior but positively to prosocial behavior. This raises the question: What happens with the predictive power of personality traits such as Honesty-Humility if prosociality requires dishonesty, as in the case of unethical loyalty? The current preregistered (https://aspredicted.org/ZSX_BS7) study aimed to tackle this question by replicating and extending prior research on the link between Honesty-Humility and unethical loyalty. To this end, we developed a new online version of a behavioral game paradigm measuring unethical loyalty in an incentive-compatible way and, in addition to self-reported Honesty-Humility levels, measured individuals' perceptions of and reasons for considering unethical loyalty as justifiable. Our findings based on a diverse sample (N = 342) showed that unethical loyalty was highly prevalent, albeit less so than suggested by prior lab-based studies. Confirming our hypothesis, Honesty-Humility showed a medium-sized negative relation with unethical loyalty, which was in part attributable to individual differences in perceiving unethical loyalty as (un)justifiable. Lastly, lower levels of Honesty-Humility were associated with perceiving unethical loyalty as justifiable particularly because it profits oneself, whereas considering unethical loyalty as profitable for the original transgressor – and thus prosocial – was not significantly related to Honesty-Humility. To conclude, the current study provides novel insights into the dispositional basis of as well as the cognitive and motivational mechanism involved in unethical loyalty. As such, we contribute to the knowledge on the relation between personality and unethical behavior in situations where dishonesty conflicts with prosociality. Next to our theoretical contributions, we provide a new online paradigm for easy and convenient measurement of incentivized unethical loyalty behavior in future research.

Session 1: Personality | Room 1

Tuesday July 2, 2024 11:30 - 11:45

**Cooperating when the end is near: the impact of cognitive ability and task complexity**

Maria Bigoni | Simona Cicognani

University of Bologna, Department of Economics | Leiden University, Department of Economics

Finitely repeated cooperation is central to many critical economic decisions. Prior evidence on finitely repeated games shows irregularities in cooperation (Embrey et al., 2017) and that more intelligent individuals cooperate more in games in which cooperating is an equilibrium (Proto et al., 2019).

This led us to investigate whether cooperation hinges on understanding and mastering the rules of a particular decision environment. Varying the focus, comprehension, and complexity involved in the decision process raises a few questions related to the choice to cooperate: is the emergence of cooperation affected by these factors? Does it also depend on how cognitively able people are?

To answer these questions, we study the relationship between cognitive ability, task complexity, and cooperation in finitely repeated social dilemmas. We run a laboratory experiment in which we borrow the design of Embrey et al. (2017) and add three main novelties: i) the comparison between two different stage games, which differ in complexity; ii) the existence of incentivized control questions; iii) a cognitive reflection test (Frederick, 2005) and a personality questionnaire (Ashton and Lee, 2009).

The experiment has a 2X2 factorial between-subject design. The first treatment variable is the stage game played. In the Prisoner's dilemma -PD condition, participants face a finitely repeated Prisoner's dilemma, whereas in the Traveler's dilemma -TD condition, they face a finitely repeated Traveler's dilemma. The implementation of incentivized control questions represents the second treatment variable.

Our results indicate that cooperation increases with experience, with no evidence of unraveling even in the last super games when subjects have gained substantial experience with the game. Regarding treatment effects, we observe that cooperation is easier to achieve and sustain in the Prisoner's Dilemma than in the cognitively more demanding Travelers' Dilemma. This evidence is robustly valid across all rounds and super games. The incentivization of Control Questions seems to produce a positive short-term effect on cooperation in the early super games played. Furthermore, we find that subjects with higher cognitive abilities (captured by the Cognitive Reflection Test) and subjects with a better understanding of the instructions tend to cooperate more.

Our experiment sheds novel light on the interplay between the evolution of cooperation and the cognitive sphere of individual decision-making. Our analysis shows that different degrees of cognitive ability and focus matter when making cooperative decisions, which makes the results relevant for making profitable cooperation in heterogeneous groups.

Session 1: Personality | Room 1

Tuesday July 2, 2024 11:45 - 12:00

**The dispositional basis of human prosociality: One common core or multiple related factors?**

Isabel Thielmann | Ingo Zettler | Benjamin E. Hilbig | Morten Moshagen

Max Planck Institute for the Study of Crime, Security and Law | University of Copenhagen | RPTU Kaiserslautern-Landau | Ulm University

Research in personality psychology consistently shows that there is a common dispositional basis underlying all socially aversive traits, such as the Dark Triad components (i.e., narcissism, Machiavellianism, and psychopathy), sadism, and exploitativeness – to name a few. This "common core" of socially aversive traits is prominently referred to as the Dark Factor of Personality (D), which is a powerful predictor of various anti- vs. prosocial tendencies and behaviors, such as cheating, cooperation, and Social Value Orientation (SVO). Although some have suggested that there may be a similar common core underlying all prosocial tendencies, empirical tests of this proposition are missing. We aim to close this gap with the current project.

In a first test, we used data from the Prosocial Personality Project (PPP), a large-scale multi-wave study measuring various traits capturing individual differences in prosocial tendencies in a demographically diverse German sample (N = 4,585). Two external expert raters determined for all traits in the PPP whether they are conceptually linked to prosociality, broadly defined. This resulted in 13 constructs measured by 16 scales that we used for the current analyses, together with measures of D, basic personality traits (HEXACO), and prosocial behavior.

Applying exploratory factor analyses and structural equation modeling, a five-factor solution provided the best fit and most interpretable factors. Specifically, the factors captured individual differences in (i) affection for others (i.e., feeling and caring for others and their well-being), (ii) dependability (i.e., being reliable and reciprocal), (iii) injunctive prosocial norms (i.e., do no harm and treat everyone equally), (iv) forgiveness (i.e., being forgiving and tolerant), and (v) trust (i.e., being trustful and believing in others' trustworthiness). Although these factors shared meaningful variance with one another, they were less strongly saturated within a common core as socially aversive traits usually are. Moreover, linking the common core of prosocial traits to D showed that the two do not represent opposite poles of the same continuum. In turn, relations with the HEXACO dimensions suggest that different aspects of human prosociality are captured by different basic personality traits. Finally, we found the common core and the five factors to differently relate to SVO and prosocial behavior across six economic games.

Overall, our results suggest that prosociality can be expressed in different ways, as reflected in the multi-faceted dispositional basis of human prosociality. We are currently in the process of pre-registering a replication study to test the robustness of these initial findings.

Session 1: Personality | Room 1

Tuesday July 2, 2024 12:00 - 12:15

**Who helps whom? An investigation of individual differences in selective prosociality**

Busra Yelbuz | Isabel Thielmann

Max Planck Institute for the study of Crime, Security, and Law | Max Planck Institute for the study of Crime, Security, and Law

Evidence consistently shows that individuals are selective in who they behave prosocial towards, helping some but not others. However, no work has directly investigated whether people systematically differ in their selective prosociality, that is, whether personality traits account for individual differences in selective prosociality. To address this, we investigated to what extent a set of personality traits – that is, the HEXACO dimensions, the Dark Factor of Personality, Social Dominance Orientation (SDO), Right-Wing Authoritarianism (RWA), and Fairness Concerns – are related to selective prosociality. In a pre-registered study, we measured selective prosociality via an incentivized decision-making paradigm where participants (N = 707) could share money with 10 recipients from different countries. Selective prosociality was operationalized as the variance in amount shared across the 10 recipients: The higher the variance, the more selective a person is in their prosociality. Besides personality traits, we also measured general attitudes towards and perceived wealth of the recipients' countries to check whether selectivity was based on differences in general attitudes or perceived wealth. Overall, we found that indeed 44% of participants were selectively prosocial (i.e., sharing money unequally across recipients), whereas 28% each were universally selfish (i.e., not sharing anything with anyone) or universally prosocial (i.e., sharing equally with everyone). Most importantly, as predicted, selective prosociality was positively related to the Dark Factor of Personality and RWA, and negatively to Openness. In addition to our hypotheses, Agreeableness was negatively related to selective prosociality. Furthermore, correlations of personality traits differed between selective prosociality (i.e., variance in amount shared) and general prosociality (i.e., overall amount shared). For example, Honesty-Humility was positively related to general prosociality but not related to selective prosociality, whereas RWA was positively related to selective prosociality but not related to general prosociality. Overall, our study provides initial evidence for the personality basis of selective prosociality and further suggests that selective prosociality should be differentiated from general prosociality. Besides the personality correlates, general attitudes towards individuals from the countries was related to amount shared with each recipient, whereas perceived wealth of recipients' countries was unrelated. This suggests selective prosociality was driven by differences in attitudes rather than perceived wealth. Overall, the study marks an initial step in investigating systematic individual differences in selective prosociality and provides important insights into its relevant trait-level drivers. To test the robustness of these findings, we are currently collecting additional data in a pre-registered study using a modified decision-making paradigm.

Session 1: Personality | Room 1

Tuesday July 2, 2024 12:15 - 12:30

**Revisiting Situational Strength: Do Strong Situations Restrict Variance in (Cooperative) Behaviors?**

Ranran Li | Daniel Balliet | Isabel Thielmann | Reinout de Vries

Vrije Universiteit Amsterdam | Vrije Universiteit Amsterdam | Max Planck Institute for the Study of Crime, Security and Law, Freiburg | Vrije Universiteit Amsterdam

Rationale: Many studies have examined the dispositional and/or contextual determinants of prosocial behaviors in social dilemmas via economic games. In addition to interindividual variability due to personality, prosocial tendencies also have been found to exhibit substantial intergroup variability across treatment groups. Yet, it remains unclear what accounts for this variability in behaviors across treatments. We address this by referring to situational strength (Mischel, 1977), according to which framework strong situations restrict variance in behavior and thus prevent personality from being expressed. In this study, we conclusively test the restricted variance hypothesis using social dilemma games as the testbed.

Method: We conducted a preregistered meta-analysis (k = 301, N = 25,670) in the context of cooperative behavior observed within the standard social dilemma paradigm. Economic game data was extracted from the Cooperation Databank (Spadaro et al., 2020), where standardized and comparable measures of prosocial behaviors were annotated.

We first meta-analytically compared the variability between groups by examining whether situations hypothesized to be stronger based on theoretical grounds (e.g., presence vs. absence of punishment, reward, low vs. high anonymity) indeed resulted in lower behavioral variance. Second, we collected perceived situational strength ratings in specific experimental scenarios from independent raters and examined if increases in perceived situational strength were associated with decreases in behavioral variance across experimental conditions.

Results: We found that strong, compared to weak, situations (theorized and validated via perception ratings) indeed restricted behavior variance. To be more specific, mostly, situations hypothesized to be stronger based on theoretical grounds (presence of punishment or reward, low anonymity, low uncertainty, etc.) indeed yielded lower variance in cooperation. Moreover, ratings on perceived situational strength of specific experimental conditions (k = 138, n_studies = 41) further supported the hypothesis that higher levels of perceived situational strength were associated with less variance in behavior.

Conclusions: Meta-analytic comparisons of behavioral variation in economic game data gave support to the situational strength hypothesis, such that situations hypothesized to be stronger indeed produced restricted variance in cooperation. Thus, variability in (prosocial) behaviors across situations can be accounted for by situational strength. The current study corroborates the situational strength hypothesis with an empirical basis.

Our findings also have important theoretical implications for understanding the situational forces shaping social behavior (e.g., reinforcement learning, social norms and reputational concerns) and advancing research on person-situation interactions.

Session 2: Evolution of cooperation | Room 2

Tuesday July 2, 2024 11:00 - 11:15

**The influence of similarity perceptions on the evolution of cooperation and confrontation**

Ilan Fischer | Lior Savranevski

University of Haifa | University of Haifa

By assuring aversive actions are followed by similarly aversive reactions, legislators of antiquity aimed to reduce belligerence and motivate cooperation. In fact, action-reaction similarity has been practiced throughout human history, typically manifested by laws of retributive justice (lex talionis), where offenses are followed by identical punishments, inflicted by the central authority. Retributive justice was practiced in Mesopotamia, the Arabian Peninsula and the Mediterranean region. It was embedded in the laws of Eshnunna (2000 BCE), the code of Hammurabi (1750 BCE), the Bible (Exodus 21:23–27) and the Quran (Surah 2, 178).

In the talk we will present experiments, involving over five hundred participants, showing how similarity perceptions drive cooperation and confrontation across several strategic decision types. Examining the choices made in three one-shot symmetric conflict games: the prisoner's dilemma, the chicken, and the battle of the sexes, we show how a short encounter accounts for the formation of subjective similarity perceptions, which together with the expected payoffs of the game determine the choice of the preferred alternative. We will then generalize the experimental findings and describe the role of similarity perceptions for all two-by-two games, specifically for a subset of fifty-seven games that are sensitive to similarity perceptions with the opponent. We will suggest that this mechanism is key to the understanding of social dilemmas and the evolution of cooperation and confrontation, and show that it complements both kin and group selection mechanisms.

Session 2: Evolution of cooperation | Room 2

Tuesday July 2, 2024 11:15 - 11:30

**Efficiency and resilience of cooperation in asymmetric social dilemmas**

Christian Hilbe

Max Planck Institute for Evolutionary Biology

Direct reciprocity is a powerful mechanism for cooperation in social dilemmas. The very logic of reciprocity, however, seems to require that individuals are symmetric, and that everyone has the same means to influence each other's payoffs. Yet in many applications, individuals are asymmetric. Therefore, we study the effect of asymmetry in linear public good games. Individuals may differ in their endowments (their ability to contribute to a public good) and in their productivities (how effective their contributions are). Given the individuals' productivities, we ask which allocation of endowments is optimal for cooperation. To this end, we consider two notions of optimality. The first notion focuses on the resilience of cooperation. The respective endowment distribution ensures that full cooperation is feasible even under the most adverse conditions. The second notion focuses on efficiency. The corresponding endowment distribution maximizes group welfare. Using analytical methods, we fully characterize these two endowment distributions. This analysis reveals that both optimality notions favor endowment inequality: more productive players ought to get higher endowments. Yet the two notions disagree on how unequal endowments are supposed to be. A focus on resilience results in less inequality. With additional simulations, we show that the optimal endowment allocation needs to account for both the resilience and the efficiency of cooperation.

Session 2: Evolution of cooperation | Room 2

Tuesday July 2, 2024 11:30 - 11:45

**A computational model of social sanctioning for encouraging division of labor**

Anil Yaman

Vrije Universiteit Amsterdam

Division of labor allows individuals to specialize in different roles and work together to achieve greater welfare as a group. Unlike most other animals, humans learn by trial and error during their lives what role to take on. However, when there are some roles that are less attractive than others but critical for the success of the group, then a social dilemma emerges: whether to take the less attractive roles for the group success or follow one's own interest that could risk the success of the whole. Consequently, a fundamental question arises: how can division of labor emerge in groups of self-interested and lifetime-learning individuals?

We propose a computational model of social norms that makes it possible to converge on role arrangements beneficial for the whole group. In our model, the social norms are decentralized social sanctioning mechanisms. All individuals within a group follow two stages: (1) select their roles based on the payoffs, (2) impose social sanctions based on their role selections.

The social sanctioning rules are adopted by all members of the group as the norm and optimized through a cultural evolution process based on cultural group selection. During this process, variations of the norms are generated through random changes (i.e. mutation), tested and accepted if they lead to a better average group payoff. Our social sanctioning model was tested on N-player spatial games that aim to simulate real-world social dilemmas such as tragedy of the commons.

Overall, we find that the social norms that emerge as a result of the evolutionary process function as redistribution mechanisms of the payoffs where individuals periodically pay others to incentivize them to perform beneficial roles for the group that they would not otherwise select. Consequently, these mechanisms allow groups of self-interested individuals to discover effective division-of-labor arrangements through lifetime-learning. On the other hand, we observe that different evolutionary processes lead to the emergence of complex norms. Introducing complexity regularization as a fitness penalty based on the complexity of the norms promotes convergent evolution to simpler social norms in independent evolutionary processes.

Session 2: Evolution of cooperation | Room 2

Tuesday July 2, 2024 11:45 - 12:00

**The evolution of group reciprocity**

David Hugh-Jones | Moti Michaeli | Ro'i Zultan

Independent | University of Haifa | Ben-Gurion University of the Negev

Reciprocity is a powerful motivation. People behave towards others as those others behaved toward them. At times, reciprocity extends beyond such direct reciprocity to include agents who were not part of the initial interaction. Such indirect reciprocity can take two forms. Downstream reciprocity involves treating an agent according to the way in which she treated others. Upstream reciprocity refers to an agent treating another agent as she was previously treated by unrelated others. The evolutionary literature proposes mechanisms by which downstream reciprocity can evolve. Upstream reciprocity, however, is more difficult to explain in evolutionary terms.

We introduce the concept of Group reciprocity. Group reciprocity is a form of upstream reciprocity, whereby an agent reciprocates the actions of another toward other members of that other's group. Group reciprocity can explain the spreading of intergroup conflict, but also helps to contain intergroup conflict from spreading beyond group boundaries. In previous work, we showed that humans have an innate tendency to group-reciprocate. Here, we show that group reciprocity can evolve as a result of a group-level selection process.

We show that group labels are a sufficient condition for the evolution of group reciprocity. The intuition is the following. With enough group reciprocators in the population, when agent i helps an out-group member she incurs an individual cost but increases the probability that the helped agent will help members of i's group in the future. Thus, helping promotes a pure public good, namely the group reputation. Selfish types, who never help, benefit individually, but harm their group. Consequently, groups with a higher share of group reciprocators (lower share of selfish types) have on average, a higher fitness. This gives rise to an intrademic group selection model: if the benefit to cost ratio of helping is large enough, selfish types---despite having the highest fitness in their group---will have lower fitness than group reciprocators overall. We complement the theoretical analysis with simulations that relax some of the assumptions of the model.

Our results relate more broadly to the evolution of group cognition. Existing theories aiming to provide ultimate explanations for why people think of groups as entities rely on either meaningful groups or assumptions regarding human cognition that require an evolutionary explanation themselves. In contrast, we show that group cognition can evolve as a result of mere labeling of individuals as belonging to groups, even when the groups are otherwise meaningless.

Session 2: Evolution of cooperation | Room 2

Tuesday July 2, 2024 12:00 - 12:15

**The evolution of risk-aversion and human risk-taking behavior in interacting public goods**

Mohammad Salahshour Mehmandoust Olia | Nico Gradwohl | Wataru Toyokawa | Johannes Kotz | Jingyu Xi | Wolfgang Gaissmaier | Iain D. Couzin

Max Planck Institute of Animal Behavior, Center for the Advanced Study of Collective Behavior, University of Konstanz. | Center for the Advanced Study of Collective Behavior, University of Konstanz. | Center for the Advanced Study of Collective Behavior,

Rationale: We combine evolutionary models and online behavioral experiments to address human risk aversion and cooperation in public goods experiments.

Methods: We introduce an evolutionary model where individuals have a choice between certain and uncertain public goods. Individuals' institution choice is subject to reinforcement learning. Both cooperation decisions and reinforcement learning parameters are evolvable genetic traits. Furthermore, to empirically address human risk aversion and prosocial behavior in social dilemmas, we perform online behavioral experiments where groups of 5 participants in the experiment can choose between two institutions to play a public good game for 20 rounds. In each round, participants receive 10 money units and can decide how much to invest in the public good. In the control condition, both public goods are certain with an enhancement factor of 2 and in the treatment conditions one of the public goods is uncertain and its enhancement factor is drawn in the interval [0,4] uniformly at random.

Results: The evolutionary model suggests that when the average enhancement factors of the two institutions are the same, reinforcement learners evolve to prefer a certain institution over an uncertain one. Such risk aversion can be beneficial due to a higher frequency of cooperators in certain institutions. However, when the uncertain institution has a slightly higher enhancement factor, the risk aversion may evolve but can be sub-optimal as the average payoff of the uncertain institution can be higher. The evolutionary model suggests the existence of an uncertain institution is detrimental to cooperation when the average productivity of the uncertain institution is lower or equal to that of a certain institution due to the evolution of risk-averse agents. Surprisingly, uncertainty can increase cooperation when the productivity of the uncertain institution is higher than that of certain institutions due to the evolution of risk-taking strategies. The results of behavioral experiments show that humans develop an unrealistically high perception of the productivity of an uncertain situation when participants have a choice between certain and uncertain institutions with similar average enhancement factors. While unrealistic, consistently with the results of the evolutionary model, this perception induces risk-taking behavior, promotes a stronger preference for an uncertain institution, and improves human cooperation.

Conclusion: Our study sheds light on the conditions under which risk aversion or risk-taking can evolve in uncertain social dilemmas and reveals that an unrealistically optimistic perception of the productivity of public good can foster risk-taking behavior, leading to improved cooperation.

Session 2: Evolution of cooperation | Room 2

Tuesday July 2, 2024 12:15 - 12:30

**Navigating social dilemmas: adaptations to infer fitness interdependence promote the evolution of cooperation**

Marco Colnaghi

VU Amsterdam

Human cooperation likely evolved in heterogeneous social environments, characterized by large variation in fitness interdependence – the extent to which an individual's fitness is affected by others. In such heterogeneous environments, cognitive adaptations to infer interdependence could have provided a selective advantage, allowing individuals to maximize their fitness by deciding when and with whom to cooperate. A large body of experimental work indicates that humans are able to infer, and respond to, variation in interdependence; yet, to date, the link between heterogeneity, inference, and cooperation remains unclear, partly because most theoretical studies focus on a single type of social dilemma.

I introduce a game-theoretical framework to study how the ability to infer interdependence shapes cooperation in heterogeneous social ecologies, where individuals engage in multiple different social dilemmas (Prisoner's dilemma, Stag Hunt, and Snowdrift/Chicken game). I investigate the evolutionary dynamics of a population where adaptive agents, who incur a cost to infer the degree of fitness interdependence, compete against fixed-behavior agents, which always cooperate or defect. When more diverse social ecologies are considered, adaptive strategies became more stable and provide a stronger selective advantage. Heterogeneity promotes the evolution of costly cognitive inference of interdependence, and this in turn enhances cooperation in a number of social dilemmas. Using a combination of evolutionary game theory and agent-based modelling, this work links variation in fitness interdependence with the evolution of inference and cooperation, shedding new light on the connections between ecology, behavior, and cognitive evolution.

Session 3: Climate change & Health threats | Room 1

Tuesday July 2, 2024 14:00 - 14:15

**Who is going to save the planet? The relationship between Passive risk-taking, Social attitudes and Pro-environmental behaviors**

Lidor Krava-Farkovitsh | Eliya Chulpaev | Yoella Bareby-Meyer

Ben-Gurion University of the Negev | Ben-Gurion University of the Negev | Ben-Gurion University of the Negev

Study rationale: Global warming, air pollution, and the extinction of animal species are serious environmental hazards that pose a threat to humanity. Although many of these threats result from human actions and can be managed, people often fail to take the necessary preventive measures. This behavior reflects people's tendency to take passive risks (risks that are embodied in inaction). Because people are held less accountable for their inaction compared to their action, accordingly, passive risk-taking is perceived as less risky than equivalent active risk-taking. Another factor contributing to the failure to act and may worsen the diminished sense of responsibility in passive risk tendency lies in the social dilemma aspect of the problem. To engage in pro-environmental behaviors, one needs to prioritize the greater good (such as using public transportation) over personal gain (such as driving one's own car). Thus, this study aims to examine the predictive roles of passive risk-taking (compared to active risk-taking) and social attitudes in pro-environmental behaviors.

Methods: Two studies were conducted using a total of 294 MTurk participants. The first study serves to develop a pro-environmental behavior questionnaire and to assess its relation with the tendency to take passive risks as measured with the PRT scale. In the second study, participants answered the pro-environmental behaviors questionnaire, and as predictors, we measured their PRT and active risk-taking tendencies (using the DOSPERT scale). Additionally, we measured their social attitudes using the SVO scale, their environmental knowledge, and awareness.

Results: We developed a questionnaire that measures pro-environmental behavior in different domains, including recycling, purchasing green products, transportation, consumption, and energy. The scale has criterion validity, as it is correlated with the PRT scale, and also has high test-retest reliability. Hierarchical regression was conducted to identify the contribution of personal characteristics in predicting pro-environmental behaviors. As expected, environmental awareness and knowledge were found as significant predictors. Importantly, while controlling for these predictors, in line with our expectations, the tendency to reduce passive risk was a significant predictor of pro-environmental behaviors. Surprisingly, active risk-taking was positively correlated with pro-environmental behavior, and SVO was an insignificant predictor.

Conclusions: Exploring the tendency for pro-environmental behaviors from a risk-taking perspective is a novelty in the field of climate change. The results suggest that individuals inclined towards decreasing passive risk and, surprisingly those inclined towards active risks may exhibit greater sensitivity to environmental hazards. This may reflect a risk homeostasis effect, which requires further research.

Session 3: Climate change & Health threats | Room 1

Tuesday July 2, 2024 14:15 - 14:30

**The Social Dilemma of Climate Policy**

Simon Columbus | Robert Böhm

Massachusetts Institute of Technology | University of Vienna

Institutional solutions are needed to combat climate change. Yet, efficient global institutions may impose costs on local communities. This creates a social dilemma in the choice of climate policies. In this context, citizens may be opposed to globally efficient policies if they believe that other countries will not adopt similar steps to fight climate change. We use surveys and incentivised experiments using a novel policy choice paradigm to study the role of beliefs in preferences over climate policies and the ability of democracy to promote compliance with climate policies. In a cross-national survey (N = 604, Germany and US), even citizens who were concerned about the threat of climate change were reluctant to endorse local policies if they believe that citizens in other countries do not support such policies. Building on this initial evidence, we introduce a new experimental game which models the social dilemma of climate policy. In this game, group members can vote on policies which maximise the local or global efficiency of contributions to the public good. Subsequently, they can make costly contributions to protect the group against catastrophic loss of income. In two incentivised, interactive experiments (total N = 1730), we use this paradigm to study (a) support for locally costly, but globally efficient policies, (b) whether individuals are more willing to comply with globally efficient policies if they are established democratically, and (c) the effect of intergroup trust on support for globally efficient policies and policy compliance. Both social preferences and positive expectations about the out-group's policy choice were associated with support for a globally efficient policy. Democracy itself had a positive effect on first-order beliefs and personal norms but did not promote contributions to the public good overall. This can be explained by heterogeneous effects of democracy: while 'yes' voters responded positively to democratically-chosen policies, 'no' voters responded negatively. Finally, positive expectations about the out-group's contribution to the public good did not promote cooperation, but neither did they crowd out contributions to the public good. Overall, we find that democracy can promote the adoption of globally efficient policies, and that when adopted, such policies increase overall efficiency without undermining individual contributions to the public good. These results can inform the promotion and adoption of institutional solutions to combat climate change.

Session 3: Climate change & Health threats | Room 1

Tuesday July 2, 2024 14:30 - 14:45

**Climate Change Ignorance: Passive Risk Taking, Self-Efficacy, and Mitigation Intentions**

Yael Mintz | Yoella Bereby-Meyer | Tehila Kogut

Ben-Gurion University of the Negev | Ben Gurion University of the Negev | Ben Gurion University of the Negev

Climate change is one of the most critical challenges that the world is currently facing. Despite the situation's gravity, people deny this threat and are less concerned about it than they should be (Spence et al., 2012). Climate change involves long-term risk, which raises less attention and is consequently underestimated (Mrkva et al., 2020). It has been claimed that individuals perceive it as beyond their control, resulting in low self-efficacy (Muroi & Bertone, 2019). By their nature, ecological risks mainly result from inaction (i.e., passive risk), such as not reducing meat consumption, and thus are perceived as less risky. To avoid confronting it, people remain deliberately ignorant about climate change despite its clear relevance and potential benefits. This study examines the relationship between an individual's personality traits, self-efficacy, and willingness to undertake mitigation measures.

Method: Two experiments were conducted with a total of 256 participants, who completed the following questionnaires: (1) BICC- measures behavioral intentions toward climate change (2) PRT- measures the tendency for passive risk-taking behavior (Keinan & Bereby-Meyer, 2012); (3) SE –measures individuals' self-efficacy (Muroi & Bertone, 2019); and (4) IPS- measures the tendency to consume information (Ho et al., 2021).

Results: Findings from Study 1 showed that people who gather information (IPS) tend to take fewer passive risks ($b=-.394$, $p<.000$). We also found a significant effect for gender, indicating that men tend to take more passive risk ($b=-.201$, $p<.05$). Furthermore, self-efficacy significantly mediates the relation between passive risk-taking tendency and behavioral intentions to mitigate climate change ($b=-.162$, $p<.01$). This relation remains significant even after controlling for self-efficacy, indicating a partial mediation effect. Results of Study 2 replicated the regression model for both IPS and Gender ($b=-.302$, $p<.01$; $b=-.247$, $p=.05$ respectively). Self-efficacy mediates the relation between passive risk-taking and intentions to mitigate climate change, replicating Study 1's findings. ($b=-.206$, $p<.01$). This relation became insignificant after controlling for self-efficacy, indicating a full mediation effect of the individuals' SE on the relationship between PRT and BICC.

Conclusion: It is essential to identify the common personality traits of individuals who are less likely to take action to mitigate climate change despite its severity. Our research has revealed that men and those who avoid information tend to take more passive risks, resulting in lower self-efficacy and a lack of behavioral intentions toward climate change. This information should be taken into account when raising awareness and motivating action on the issue.

Session 3: Climate change & Health threats | Room 1

Tuesday July 2, 2024 14:45 - 15:00

**Can Climate Clubs Curb Climate Change? - An Experimental Study**

Sakshi Kasa Prasad | Tibor Neugebauer | Stefan Traub

Luxembourg Institute of Socio-Economic Research | University of Luxembourg | Helmut-Schmidt University, Hamburg

The Climate Club Theory (Nordhaus, 2016) seeks to address the issue of free-riding that often occurs in global climate agreements; suggesting exclusive tax benefits to countries that have strong climate policies, which are funded by countries with more lenient policies. Theoretically, it incentivizes abatement by rewarding club members and penalizes emissions by taxing non-members. In a non-linear public bad laboratory experiment, we examine two different ways of implementing club membership: the Competition Treatment, where countries must compete to have an abatement level equal to or above the average to join the club, and the Voting Treatment, where the threshold for membership is determined by a vote among current club members. Both treatments make the exclusionary benefits of the club salient; the voting treatment provides an additional focal point for cooperation.

The experiment was finitely repeated, had a between-subject design, and was conducted online with 145 participants using Z-tree Unleashed with students from the University of York. Using the Mann-Whitney U-test, our results confirm our pre-registered hypothesis that the club treatment leads to a significantly lower level of emissions than the baseline in the Competition treatment, however it rejects the hypothesis in the case of the Voting treatment. In the Competition treatment, the average investment in abatement declines over time as participants fail to coordinate on a stable average and eventually resort to freeriding. In the Voting treatment, the average abatement remains stable and there is an inverse relation between the threshold for membership and the number of club members. Club members vote to drive up the threshold for club entry above the predicted Nash Equilibrium, however, the non-members freeride on the club's contributions, driving down the average. In the real world, this amounts to a few countries setting climate goals too ambitious for others to comply with, thus dis-incentivizing any action on their part. While the climate club in our setting does not provide sustained high levels of abatement, it provides useful insights for policymakers.

Session 3: Climate change & Health threats | Room 1

Tuesday July 2, 2024 15:00 - 15:15

**Climate extreme events are enforced by extortionate free-riders**

Manfred Milinski | Stefania Innocenti

Max Planck Institute for Evolutionary Biology, Germany | Smith School of Enterprise and the Environment, Oxford

Will climate extreme events exert a disciplining effect increasing mitigation efforts?

In our control game, players in groups of 6 make voluntary contributions towards a target to avert a simulated climate catastrophe. With a starting endowment of £40, each player can allocate either 0, £2, or £4 in each of 10 rounds to reach a group target of £120. If the group fails, players forfeit their remaining private endowment.

In two additional treatments, groups are exposed to an intermediate climate event every second round of the game. In ""Constant fee"" treatment, if the group fails to reach an intermediate target, each player pays a penalty of £1, risking up to £4 in total. In ""Increasing fee"" treatment, each group member pays a penalty of £1 in round 2, increasing to £16 in round 8 if the group misses the respective target, generating a potential loss of up to £29.

Averting intermediate targets has no costs, nonetheless in the ""Increasing fee"" treatment by missing all targets the final goal becomes unattainable. 30 groups participated in each treatment.

Most groups did not avoid intermediate penalties, decreasing individuals' private endowment, especially in the ""Increasing fee"" treatment. Surprisingly, the percentage of groups reaching the final target was almost identical across treatments. This situation presents a paradox, also in control, where climate events were absent and cannot have exerted a disciplining effect.

In social dilemmas, some individuals act as extortionists (Press & Dyson 2012), adopting steadfast selfish strategies, thereby forcing fair players to cooperate. To attain the final target and keep their money, fair players have no choice but to try to compensate the deficit caused by their extortionate co-players.

We find that the average number of extortioners in a group did not differ across treatments, indicating that their presence is exogenous to treatment assignment.

With more extortioners in a group, the total deficit they create grows, leading fair players to contribute more. This highlights that fair players were compelled to compensate for the extortioners' deficit and did so in all treatments, generating significant gains for extortionists.

Therefore, failing to reach intermediate targets was just the by-product of the extortion process while attempting to reach the final target.

This reflects reality. Although climate extreme events dominate the news, people often ignore them hoping that others will atone for their misdeeds, as exemplified by record numbers of people traveling to holiday destinations in 2023.

Session 3: Climate change & Health threats | Room 1

Tuesday July 2, 2024 15:15 - 15:30

**Getting vaccinated for pro-social reasons: when does it work?**

Mitchell Matthijssen | Florian van Leeuwen | Mariëlle Cloin | Ien van de Goor | Peter Achterberg

Tilburg University | Tilburg University | Tilburg University | Tilburg University | Tilburg University

Rationale: If people would act only according to their self-interest, eradication of infectious diseases with vaccines would be impossible. This is because vaccination is a social dilemma. When deciding to get vaccinated, the best strategy on the individual level (e.g., not getting vaccinated because the discomfort of doing so outweighs the risk of contracting a low-risk disease) can be misaligned with the best strategy on the collective level (e.g., maintaining high vaccination coverage levels for the protection of vulnerable individuals). Therefore, previous work has proposed that interventions can increase vaccination uptake by remind people about the collective interest of vaccination and call upon their prosocial motivations to help protect others. Previous research is supportive but has two limitations. First, they exclusively measured decisions for hypothetical diseases and did not include existing (real) diseases. Second, previous studies often only considered disease characteristics (e.g., contagiousness, vulnerability) and did not include multiple characteristics at the same time. To address these limitations and test under what conditions prosociality influences vaccination intentions, we tested the following hypotheses: Prosociality is positively associated with vaccination intentions, this effect is stronger for hypothetical diseases and this cannot be explained by individual differences in social desirability, when people are less vulnerable to contract the disease, and when the disease is more contagious.

Method: We have conducted a pre-registered cross-sectional study with a representative sample in the Netherlands (N = 2355). Participants completed individual difference measures of prosociality (i.e., agreeableness), perceived vulnerability, and social desirability. We measured vaccination intentions for 9 diseases (e.g., COVID-19, Holtosis, tetanus). We varied the diseases both in type (i.e., hypothetical vs existing diseases) and contagiousness (e.g., not contagious to very contagious). Currently we are collecting data for an experimental test of the hypotheses (N ± 2200).

Results: We found that agreeableness was associated with vaccination intentions in general, was more strongly associated with existing disease, and this could not be explained by individual differences in social desirability. Furthermore, agreeableness was not more strongly associated with contagious diseases and when people were less likely to contract diseases.

Conclusion: This study showed support that studies focusing on hypothetical diseases generalize to existing diseases but did not support that these prosocial tendencies had stronger effect for other disease characteristics. This might show that people get vaccinated to protect other people irrespective of whether this is best strategy for them individually.

Session 4: Intergroup conflict | Room 2

Tuesday July 2, 2024 14:00 - 14:15

**Individual exploitation in intergroup conflicts**

Qinyu Xiao | Simon Columbus | Robert Böhm

Department of Occupational, Economic, and Social Psychology, University of Vienna | Massachusetts Institute of Technology | Department of Occupational, Economic, and Social Psychology, University of Vienna; Department of Psychology, University of Copenhag

Some intergroup conflicts start with an act of aggression perpetrated by one individual, or a small coalition, in pursuit of their own self-interest. Raids on out-groups may provide participants with material and status benefits. However, such raids can escalate into spirals of retributive violence. Thus, the pursuit of individual self-interest through the exploitation of out-groups can impose high costs on in-group members who do not share in the gains but nonetheless suffer from retaliation. This trade-off between individual gains from the exploitation of out-groups and collective costs from defending against retaliation has largely been overlooked in the experimental literature, which has often assumed that in-group members share in the spoils of war. We therefore introduce a variant of the Intergroup Prisoner's Dilemma—Maximizing Difference (IPD-MD) game which allows for individual exploitation of out-groups and collective retaliation. In a pre-registered and incentivized real-time interactive experiment (n = 376), we compared behaviors in a repeated, sequential IPD-MD game and an otherwise equal game that additionally allows participants to individually exploit the outgroup (the IPD-MD+IE game). Consistent with previous findings, in the IPD-MD, out-group harm was infrequent and did not lead to retaliatory aggression. In contrast, in the IPD-MD+IE, even initially low levels of individual exploitation quickly increased over rounds and crowded out weakly parochial cooperation, especially in groups with a highly aggressive minority. Participants anticipated the cost of retaliatory aggression: they engaged in less individual exploitation at the beginning of the repeated game than participants playing a comparable one-shot game. In line with the ethnographic literature, this suggests that groups have an interest in sanctioning and suppressing individual exploitation of out-group members. In a preregistered follow-up experiment, we test whether peer punishment can suppress individual exploitation and promote peaceful intergroup relations. Our work thus contributes to recent literature on the evolution of peace and the emergence of social norms and formal institutions to promote peaceful intergroup relations.

Session 4: Intergroup conflict | Room 2

Tuesday July 2, 2024 14:15 - 14:30

**To Share or to Freeride? Public Goods in Heterogeneous Societies**

Dora Simunovic

Constructor University

Cooperation in heterogeneous societies can be difficult to sustain. Despite the strong theoretical expectation that homogeneous (as opposed to heterogeneous) group membership and composition would enhance cooperation, while the presence of distinct minority and majority groups would reduce it, relatively few studies have tested these ideas experimentally. In the current study, participants played a public goods game as unrelated strangers, as members of a homogeneous group, or as members of either the minority or majority part of a heterogeneous group. We predicted homogeneous group membership would enhance cooperation beyond groups of unrelated strangers, but this positive effect would disappear when the group is heterogeneous. The reason for this would be contrasting expectations about minority and majority group members' contributions to the common resource in the heterogeneous condition. Two experimental studies were conducted to test these ideas.

Study 1 found no differences in cooperation across conditions. In other words, mere heterogeneity did not seem to decrease cooperation. However, predictions about cooperation by minority versus majority groups did indeed follow the predicted pattern: not only did majority group members expect significantly lower contributions from minority in comparison to the majority group, but minority members seemed to agree, likewise expecting lower contributions from fellow minority group members. This finding violates the ingroup favoritism principle and presents us with ingroup derogation.

Study 2 dug deeper into why the loss of common resources which is recorded in some heterogeneous groups outside of the laboratory did not translate into a loss of cooperation in the experimental condition. In particular, we wondered about the impact of the Shadow of the Future on the willingness to cooperate in heterogeneous versus homogeneous groups, making opposing prediction for how it would impact minority and majority group members in particular. On the one hand, expecting future interaction with the same group members should enhance cooperation across all conditions. On the other hand, the ability of minority group members (in particular) to exchange resources in a subsequent interaction should make freeriding more tempting to minority group members, but more threatening to majority group members. Results are discussed with respect to theoretical and practical implications for common resource management.

Session 4: Intergroup conflict | Room 2

Tuesday July 2, 2024 14:30 - 14:45

**Are cooperators more likely to attack out-group members than defectors? – The second experiment**

Nobuyuki Takahashi | Nobuhiro Mifune | Yoshie Matsumoto | Toko Kiyonari | Dora Simunovic | Toshio Yamagishi

Hokkaido University | Kochi University of Technology | Shukutoku University | Aoyama Gakuin University | University of Bremen

Homo sapiens is considered the only species that are capable of large-scale cooperation. One proposed explanation is the parochial altruism (PA) hypothesis, which argues that within-group cooperation and out-group aggression have co-evolved (Bowles & Gintis, 2011; Bowles, 2008; Choi & Bowles, 2007). To empirically examine the relationship between within-group cooperation and out-group aggression, we conducted the first experiment (Matsumoto et al., 2019). Undergraduate students were divided into two groups and played a one-shot public goods game (i.e., the measure of in-group cooperation). They also played a preemptive strike game (PSG) (Simunovic et al., 2013) in which they decided whether or not to push a button within 30 seconds. The group in which more participants pushed the button won and imposed a penalty on the other group. Thus, pushing the button is the measure of out-group aggression. The results showed that cooperators were no more likely than defectors to attack the other group, which provides negative support for the PA hypothesis.

The current study is Experiment 2, which further explored whether the same pattern is found in the other population of participants and the other game. Yamagishi and his colleagues have conducted a series of experiments to measure various behavioral and psychological tendencies with nonstudent samples (from the 20s to 60s) who lived around Tokyo. From the pool, we have recruited 295 participants. In addition to the inter-group PSG, participants played the IPD-MD game (Halevy et al., 2008). In this game, participants decided how to allocate their endowment. Unlike the ordinary SD game, participants had three options. Contribution to the within-group pool is the measure of in-group cooperation, while contribution to the between-group pool is the measure of out-group aggression. Withholding contribution is the measure of defection. Since they had already played various games months before, we have multiple measures of in-group cooperation.

The results again showed no relationship between within-group cooperation and out-group aggression in PSG. This pattern holds when, as the measure of within-group cooperation, we used the cooperation rate in the SD game, the overall cooperativeness made from 5 games in the previous experiments, and SVO. The results of the IPD-MD game confirmed the pattern. There was no relationship between within-group cooperation and the amount of contribution to the between-group pool. Therefore, the results of the second experiment are consistent with those of the first experiment, which cast serious doubt on the co-evolutionary model of parochial altruism.

Session 4: Intergroup conflict | Room 2

Tuesday July 2, 2024 14:45 - 15:00

**Do the Powerful Compete or Cooperate? An Intergroup Exploration of Power**

Tamar Kugler | Zixu Zhang | Sarah Doyle | Poonam Arora

University of Arizona | University of Arizona | University of Arizona | Quinnipiac University

Objective: Do the powerful compete or cooperate? Existing evidence shows that the experience of power decreases cooperative behavior (Galinsky et al., 2003; Blader et al., 2016). Yet, others demonstrate that power can lead to increased cooperation (Anderson et al., 2012; Molho et al., 2019). We argue that the causal relationship between power and cooperation is under-specified because power can be experienced at the individual level (one group member is more powerful than others) and at the group level (one group possesses more power than others); similarly, cooperative behavior can be targeted at both the ingroup and outgroup. We therefore study the relationship between power and cooperation within an intergroup setting.

Method: In four pre-registered, incentivized experiments, we manipulated individual power (high, low, neutral) through random role assignments (managers, associates, or no role; Galinsky et al., 2003). Participants then allocated 10 chips in a classic social dilemma, an intergroup competition (Kugler & Bornstein, 2013), an IPD-MD game (Halevy et al., 2012) and an IPUC game (Aaldering & Böhm, 2020). Depending on the study, their allocations generated income for themselves, their group, the collective, or reduced outgroup income.

Results: Study 1 (N = 265), a single-group social dilemma showed that the powerful behaved more cooperatively, and reported greater identification with their ingroup mediating power's effect on intragroup cooperation. Study 2 (N = 267), an inter-group competition where participants could make a costly contribution benefiting the ingroup but hurting the outgroup, revealed no difference in the allocations of the powerful and powerless, but the powerful did report greater ingroup identification, mediating the effect of power on inter-group competition. Study 3 (N = 260), where allocations could benefit the ingroup or hurt the outgroup showed no difference between the powerful and powerless individuals. Study 4 (N = 398) showed that the powerful contributed significantly less to the collective pool benefiting both the ingroup and outgroup.

Conclusions: Initial findings support the concept of "benevolent power" and demonstrate mediating the role of group identification. However, this increased cooperation is constrained within the group as the powerful did not contribute towards the collective good, thus establishing a boundary for the effect of power on cooperative and competitive behavior.

Session 4: Intergroup conflict | Room 2

Tuesday July 2, 2024 15:00 - 15:15

**Unpredictable Futures, Parochial Pro-Sociality, and Intergroup Conflict**

Zsombor Méder | Jörg Gross | Carsten K.W. De Dreu | Lennart Reddmann

Faculty of Economics and Business, Rijksuniversiteit Groningen, Groningen, Netherlands | Social and Economic Psychology, University of Zurich, Zurich, Switzerland | Social, Economic and Organisational Psychology, Leiden University, Leiden, Netherlands | M

Rationale: Groups experience carrying-capacity stress when returns from local club goods become unpredictable and may not sustain group survival and prosperity. Although we know that humans dislike uncertainty and find unpredictability stressful, how groups respond to unpredictable futures is unclear: A small body of work has found that carrying-capacity stress associates with increased group solidarity and parochial cooperation, but also with competition and intergroup conflict. Our research aims to reconcile these seemingly disparate findings in an experiment where individuals within groups had the option to contribute to local club goods with (un)predictable returns, or to engage in intergroup conflict and aggression.

Methods: We conducted a controlled, fully incentivized laboratory experiment, with participants grouped into 3-vs-3 attacker and defender teams. Each group was faced with the option to contribute to a local club good with either predictable or unpredictable returns. Simultaneously, participants engaged in an intergroup contest, investing resources in either out-group attack or in-group defense. Depending on condition, the club good pool would either yield predictable or unpredictable returns, allowing us to study the effects on club good contributions, conflict investments and ensuing dynamics of conflict.

Results: Results indicated a clear shift in group behavior under uncertainty. With unpredictable returns from local club goods, participants in attacker groups decreased their contributions and increased investments in out-group attacks. Conversely, defender groups, facing heightened aggression, invested more in in-group defense and less in their club goods. This shift in strategy led to a marked decrease in overall social welfare. Furthermore, unpredictability led both attackers and defenders to reduce coordination on their club goods and increase free-riding, while prompting attackers to enhance cooperation and coordination in out-group aggression.

Conclusions: Our findings suggest that carrying-capacity stress, induced by environmental unpredictability, prompts a shift from cooperative engagement in local club goods to outgroup aggression and intergroup conflict, diminishing overall social welfare. This research contributes to the existing macro-level literature on environmental deterioration and volatility – key factors underlying carrying-capacity stress – and heightened prevalence of intergroup conflict and violence, helping to understand the behavioral underpinnings of this link. In this context, our findings indicate that geo-political or ecological volatility may play an important role in intergroup conflict and suggest that helping groups to reduce or mitigate such stressors can prevent intergroup conflict and violence.

Session 4: Intergroup conflict | Room 2

Tuesday July 2, 2024 15:15 - 15:30

**Leader Rhetoric Escalates Intergroup Conflict**

Luuk Snijder | Jörg Gross | Carsten De Dreu

Leiden University | University of Zurich | Leiden University

Unprovoked attacks on out-groups are illegal by international law, may be seen as less morally defensible than defending against such provocations, and free-riding is more often observed during attack rather than defence. Leaders who can benefit from winning intergroup conflict may thus be tempted to portray their desire to attack outsiders as defending against out-group hostilities. While fitting anecdotal evidence from historical cases, a systematic analysis of leader rhetoric during attack and defence is missing, and we have no insight in whether and how portraying attack as defence impacts conflict dynamics and outcomes.

We addressed these questions with an archival analysis and two incentivised experiments. Study 1, an archival analysis of 261 war manifestos between 1508 and 1945, showed that unprovoked attacks often contained clear yet untrue reference to the country's need to defend its values and territories against outsider intrusions and hostilities. Study 2 asked participants on Prolific (N = 164) to classify excerpts of leader speeches delivered at the onset of international conflicts as attack or defence motivated. Results showed that participants more accurately classified leader speeches from defender countries as 'defence' (70.4%, versus 29.6% as attack), than speeches from attack countries as 'attack' (46.3%, versus 53.7% as defence). Study 3 studied the emergence and consequences of leader rhetoric in the behavioural laboratory (N = 312). Participants, two 'fighters' and one 'leader', engaged in an intergroup attacker-defender contest. On each round, leaders were informed if their group was in the attacker or defender role and had to convey this information to their fighters. In the control condition, this information always had to be truthful, but in the deception condition this information could be truthful or not. Fighters were informed of the role their leader had communicated to them and could contribute their resources to conflict. We found that during attack leaders deceived their fighters on 38.85% of rounds (communicating to its fighters that they were in defence), whereas during defence they rarely deceived their fighters (13.27%). Because fighters who believed they were defending contributed more to conflict than those believing they were attacking, leader rhetoric escalated intergroup conflict and increased its economic waste.

Session 5: Sustainability | Room 1

Wednesday July 3, 2024 09:00 - 09:15

**Who, how much and when: effect of endogenously suggested rules and inequality on a common resource game**

Giulia Priolo | Laila Nockur

Aarhus University, Department of Psychology and Behavioural Sciences | Aarhus University, Department of Psychology and Behavioural Sciences

Fair management of common resources is pivotal for achieving sustainability, but entails a social dilemma between personal, short-term interests and collective, long-term ones. Moreover, inequality can complicate the issue further by reducing cooperation (Suchon & Theroudé, 2022). Introducing democratic institutions, (e.g., voting) can alleviate this detrimental effect (Nockur et al., 2020; 2022), although results are mixed. A deeper understanding of inequality's impacts on cooperation and possible solutions, not only provides valuable theoretical insight but also informs real-life strategies for enhancing sustainability.

This pre-registered study investigates whether endogenously suggested rules and inequality impact a common resource game's (renewable pool, ten rounds, four players) dynamics. Specifically, we examine whether the possibility to endogenously determine the maximum extraction levels (I.e., MELs) of group members influences sustainable resource management and how this depends on MELs' initial distribution. We further seek to explore suggestions' characteristics and voting behavior.

The study entails a 2 (Voting condition: Vote vs No Vote) x 3 (Type of players: Advantaged vs Disadvantaged player (in unequal groups) vs. Equal players) between-subject design. 460 Participants are randomly assigned to an Inequality condition and either play in a group in which all players have the same MELs (25%), or in a group in which two players have a high MEL (33%) and two players have a low MEL (17%). Participants in the Vote condition can suggest new MELs distributions and vote on suggestions after each round, whereas this option is not available in the No Vote condition.

Results suggest that the Voting condition does not enhance sustainability but reduces the disparity in the average percentage extracted between Advantaged and Disadvantaged players. No significant differences are observed in the average suggested MELs, which tend to be unsustainable across conditions and player types. However, Unequal groups exhibit higher variance in suggested MELs (vs. Equal, $p = .003$, $d = 0.69$) and generate more suggestions (vs. Equal, $p = .029$, $d = 0.30$). Furthermore, Disadvantaged players tend to be the first to suggest ($p = .057$, $OR = 1.98$) and make their suggestions earlier ($p = .061$, Cohen's $d = 0.38$). Voting dynamics (e.g., likelihood of voting for a suggestion) were also assessed both at Condition and Type of player levels.

Our findings indicate that endogenously suggested rules may not improve sustainability, but can contribute to reducing inequality between Advantaged and Disadvantaged players.

Session 5: Sustainability | Room 1

Wednesday July 3, 2024 09:15 - 09:30

**What makes communication effective to govern shared resources?**

Minwoo Ahn | Raksha Balakrishna | Marco Janssen | Michael Simeoni

Arizona State University | Arizona State University | Arizona State University | Arizona State University

Rationale: The tragedy of the commons narrative proposed in 1968 by biologist Garrett Hardin adopts the simple logic that people who share a common resource will overharvest it. This led to the advocating of privatization or nationalization of shared resources. Empirical research has demonstrated that communities can organize themselves to manage their shared resources under the right conditions.

Behavioral experiments in the lab and the field have found communication very effective. The effectiveness of cheap talk, in which participants can make promises that cannot be enforced, is puzzling. Proposed explanations include the development of trust within the group, coordination of actions, and improved understanding of the game. A limited number of studies have analyzed the content of communication, typically using manual coding of communication phrases. No clear insights are derived from those studies.

Methods: In this study, we combine data from 4 types of resource games (groundwater extraction, surface water irrigation, spatial foraging, and community infrastructure) in which participants use computer chat to communicate. The resulting data of 1500+ rounds and 100+ groups are analyzed using computational text analysis. We use sentiment analysis tools to measure overall positive and negative valence of communication and estimate the contribution of sentiment on decisions in the experiment.

Results: We show that positive sentiment meaningfully increases levels of cooperation. However, the positive sentiment effect does not impact each type of game to a similar extent. To unpack the contents of communication, we use Structural Topic Model (STM) to estimate topic contents and prevalence. Results suggest that coordination-related themes are dominant topics discovered across games, along with other game-specific topics. Topic estimation models suggest that coordination is a major contributor to cooperative outcomes in some games.

Conclusion: We conclude that decision context is important in shaping communication sentiment and contents, and they must be understood simultaneously to explain cooperative behavior.

Session 5: Sustainability | Room 1

Wednesday July 3, 2024 09:30 - 09:45

**Binding the Future: Far-looking altruism boosts long-term sustainability**

Eliran Halali | Oren Perez

Bar-Ilan University | Bar-Ilan University

Rationale: Intergenerational cooperation, essential for addressing some of the most pressing challenges facing humanity today, is particularly challenging to achieve. The risk of asteroid impact, biodiversity, and AI safety are archetypal examples of such intergenerational social dilemmas, but the most formidable and pressing one is climate change. A unique challenge of intergenerational social dilemmas is that key mechanisms that facilitate cooperation in single-period social dilemmas (reciprocity and third-party punishment) or compensate for its absence (formal compliance mechanisms) are lacking in the intergenerational context. An effective solution for fostering multigenerational collaboration could involve the implementation of a commitment mechanism (CM) imposed by the current generation on future generations, compelling them to continue and collaborate with subsequent generations.

Methods: We experimentally examined the behavioral aspects of implementing a CM to enhance intergenerational collaboration. In a preregistered study with 990 participants, we investigated whether people are willing to (altruistically) invest in a costly CM and explored the effect of introducing the option to implement a CM on the sustainability of the common pool across generations. We used a modified version of an inter-generational public goods game (Hauser et al, 2014) in which participants were assigned to five-person groups (chains). Within each chain, each participant represented one of five generations and was endowed with a common pool of 100 units that replenished itself (i.e., refilled to 100 units) if 50 units were transferred to the next generation. We compared two experimental conditions: a control condition and a CM condition in which participants could invest 10 additional units in a CM that limits the choices of the following generation, ensuring their continued cooperation with the next generation.

Results: We find a widespread endorsement for using CMs, despite their associated costs. This seems to reflect ‘far-looking altruism’: the inclination of some individuals to forego personal gain to improve not just the welfare of the next generation but also that of a more distant one (the third). We also find that CMs yield long-term benefits, by increasing the sustainability rate (that is, the proportion of chains that managed to sustain the common pool across all generations). Finally, our results imply that once set in motion, CMs are highly persistent, as subsequent generations tend to continue utilizing them.

Conclusions: A CM could be an effective solution for fostering multi-generational collaboration. Its implementation, imposed by the current generation on future generations, compels them to continue and collaborate with subsequent generations. These results have important policy implications.

Session 5: Sustainability | Room 1

Wednesday July 3, 2024 09:45 - 10:00

**Over-harvesting like there is no tomorrow: Evidence from dilemma-free common pool games**

Jan K. Woike | Sebastian Hafenbrädl

University of Plymouth, UK | IESE Business School

Why do people over-harvest renewable resources and underestimate spiraling debts? In previous studies, we observed excessive harvesting even in variants of common-pool resource games designed to eliminate the inherent social dilemma. An exponentially growing resource with a high growth factor made non-harvesting in early rounds the individually and collectively optimal strategy. Nonetheless, participants reduced their payoffs by harvesting substantial amounts too early. Here, we explore whether this behavior can be explained by a perceived competitive structure or by a lack of cognitive insight.

In two new studies, we study behavior in one-player variants of common-pool resource games with optimal strategies and exponentially growing resources. In Study 1 (n=243), we demonstrate that the majority of participants over-harvest the resource at a personal cost. Costly over-harvesting increases when the resource is additionally diminished by forces outside of the participant's control. At the same time, this does not depend on whether the influence is due to a competitor or due to natural waste. In Study 2 (N=699), we again observe over-harvesting irrespective of the presence or absence of waste or a positive or negative (debt) framing of the game. We demonstrate relationships between game behavior and predictors including financial literacy, numeracy, and exponential reasoning.

We conclude that over-harvesting is unlikely to be eliminated without addressing the cognitive challenge of anticipating exponential growth and adjusting one's strategy accordingly.

Session 5: Sustainability | Room 1

Wednesday July 3, 2024 10:00 - 10:15

**The Effect of Ecolabels and Attention on Producer's Sustainable Decision-Making**

Jan Hausfeld | Julian Kirschner | Jan Engelmann | Natalie Lee

University of Amsterdam | University of Amsterdam | University of Amsterdam | University of Amsterdam

Ecolabels, standards, and information provision can have significant systematic effects on producer and consumer behavior, potentially providing tools to shift industries towards sustainability. Yet, there is limited evidence from controlled environments or their effects on producers. We investigate the impact of ecolabels and information order on sustainable "producer decisions" within a novel common resource paradigm framed as producer decisions.

Paradigm: In a deforestation setting, participants decide how many trees to harvest from a common forest which has a steady regrowth rate of 40 trees per round. Each harvested tree yields profit across 10 rounds. Conserving trees not only benefits the commons in the game but trees remaining after 10 rounds are planted in the real world, creating a dynamic environment where environmental as well as individual behaviors are incentivized.

Between subjects we vary the ecolabel (none, symbolic, monetary) and order of the information Each participant plays an interactive (groups of 4 players) and a computerized version (3 automated group members). We ran the lab study in Netherlands and Germany (N=244) and used eye-tracking for 96 participants.

Results/Conclusions: We show that both labels reduce the initial and average amount of trees cut per round, they seem to function like a coordination device. This leads to common forests surviving longer, and environmentally as well as economically more efficient outcomes. The monetarily incentivized ecolabel does not lead to significantly different or better outcomes than a mere symbolic one, even indicating a crowding out effect towards the end of the game.

We explore the role of information order and gaze patterns through eye tracking (n=96) on behavior. We find that participants are biased towards information presented in the most salient (top left) position and that increased time spent on ecolabel information and remaining trees correlates with more sustainable decisions. We conclude that ecolabels can shape standards as well as norms, hence providing a potentially powerful tool for a shift towards a sustainable future.

Session 5: Sustainability | Room 1

Wednesday July 3, 2024 10:15 - 10:30

**Can exclusion prevent the Tragedy of the Commons? An experiment comparing excludable vs. non-excludable resources**

Erik de Kwaadsteniet | Welmer Molenmaker | Eric van Dijk

Leiden University | Leiden University | Leiden University

Background: As shown in experimental studies and real-world common resource dilemmas (e.g., over-fishing), people often over-harvest from common resources. That is, they tend to harvest so much that the resource cannot replenish itself and declines: the well-known "Tragedy of the Commons". What often characterizes such common resources is that they are non-excludable, as all individuals can freely harvest from them. In this research, we investigate whether the possibility to exclude individuals from harvesting can prevent overuse and resource depletion from occurring. Additionally, we test whether exclusion opportunities might instead lead to a so-called "Tragedy of the Anticommons", which occurs when excessive use of exclusion leads to underuse.

Method: In a preregistered experimental group study (N = 102), 3-person groups participated in a 20-round replenishable common resource dilemma task (cf. Roch and Samuelson, 1997). Groups started the first round with a common resource of 300 valuable points, from which they could freely harvest. However, if they harvested more than available in the resource, the resource was depleted. This meant that they would not earn any points that round, and the group task would be ended prematurely. In case the resource had not been depleted yet, the common resource was replenished by multiplying the points left over in the resource by 1.2 (= the replenishment rate), and a new round was started. The resource could not increase to any number higher than 300 points (300 was the maximum resource size).

There were two between-subjects conditions: a non-excludable vs. an excludable condition. In the excludable condition, group members could not only harvest points themselves, they could also block others from harvesting. That is, at the end of each round, after receiving full feedback about the harvests and resource size, they had the opportunity to block one group member from harvesting. If an individual group member was blocked by at least one other group member, this individual was not allowed to harvest in the next round.

Results and Conclusion: In both conditions, over-harvesting was frequently observed (and underuse very infrequently), and the common resource quickly decreased in size. Nevertheless, in the excludable condition participants seemed to have a strong preference not to exclude anyone (which was chosen in 67.1% of the cases). Still, in this condition resource depletion was significantly less severe than in the non-excludable condition. These findings suggest that the possibility to exclude one another can be effective in ameliorating (but not preventing) the Tragedy of the Commons.

Session 6: Trust | Room 2

Wednesday July 3, 2024 09:00 - 09:15

**Trust in Gender and Profession—A Social-Category Account**

Oliver Schilke | Tamar Kugler | Zeyu (Arthur) Xue

University of Arizona | University of Arizona | University of Arizona

Trust, the willingness to make oneself vulnerable to others (Rousseau et al., 1998; Schilke et al., 2021), is critical to the effective functioning of our society. When interacting with others, trustors may form what Foddy et al. (2009) termed "group-based trust"— use social categories of the counterpart as cues for trustworthiness. We study to what extent people rely on social category-based information to infer trustworthiness and whether reliance on such information facilitates accurate trust decisions.

We first explore how social category information about the gender or profession of the trustee affects trust. Our pre-registered experiment used a 2 (gender information: provided vs. no information) X 2 (counterpart gender: female vs. male) factorial designed. We measured trust using the rely-or-verify game (Levine & Schweitzer, 2015; Schilke & Huang, 2018), and measured trust accuracy using the expected payoff to the trustors in the different conditions.

Our findings show that female trustees (90% accurate, N = 50) were slightly more trustworthy than male trustees (83.3%, N = 48). When gender information was presented, trustors trusted female trustees (77.5%, N = 213) slightly more than male counterparts (71.8%, N = 209), and in both information conditions trust was slightly higher than in the no information conditions. Trustors' expected earnings fell just short of being higher when gender information was presented compared to no information presented ($F(1, 714) = 3.123$, $p = .078$)

The results of this first study showed that participants were generally under-trusting. When social category information (gender) was provided trust went slightly up, possibly increasing trust accuracy with it, but gender as a social category was not a strong cue regarding trustworthiness. Our second study (to be conducted over the next months and presented at the conference) will test whether profession information (e.g., salesman, politician, nurse, professor) has a bigger effect on trust and trust accuracy. We will discuss whether the stereotypes around social categories are useful or detrimental to trust judgments.

Session 6: Trust | Room 2

Wednesday July 3, 2024 09:15 - 09:30

**Emotional determinants of trust behavior**

Jan B. Engelmann | Federica Farolfi | Li-Ang Chang

Amsterdam School of Economics | Adam Smith Business School, University of Glasgow | Amsterdam School of Economics

Rationale: The social emotional and social cognitive processes involved in trust decisions are difficult to measure and disentangle in traditional economic game settings.

Methods: To address this gap, we developed a novel questionnaire to assess participants' emotional reactions to betrayal in familiar hypothetical real-world scenarios. Using this questionnaire we investigated how betrayal-reactive emotions and cognitions relate to decisions and preferences in two incentivized economic games: the trust and the dictator game.

Results: Results from our study indicate that the affective and social cognitive reactions that subjects self-reported in hypothetical real-world scenarios that involve betrayal show specific associations with trusting behavior and subjects' tendency to donate. Our main observation is that betrayal sensitivity is a robust and specific positive predictor for trust behavior. Similarly, the personal distress measure of the IRI is also positively associated with trust. Jointly, these results suggest that individuals that experience higher levels of reactive betrayal, as well as the related construct of personal distress, also exhibit a greater willingness to make themselves vulnerable to such betrayal.

In follow-up analyses we split participants into three betrayal sensitivity types using k-means clustering on the reactive betrayal score. We first test whether groups differ in psychological factors, such as impulsivity, and general social preferences, such as trustworthiness levels. We do not find strong evidence for such group differences across. However, we observe a significant group difference in empathic concern, while no other IRI items show significant effects. The group difference in empathic concern is driven by the higher empathic concern score of the medium (mean IRI EC = 19.28) and high betrayal sensitivity group (mean IRI EC = 19.9), compared to the low betrayal sensitivity group (mean IRI EC = 14.75). This suggests that the increase in empathic concern exhibited by these high betrayal sensitive types might make them more sensitive to the affective components of cooperation and non-cooperation. We furthermore assess the role of social distance in economic games and find that it robustly predicts trust and dictator game behavior.

Conclusions: We conclude that social preferences as traditionally defined in economics are not uniquely responsible in explaining trusting and giving behavior. Our findings suggest that the concept of social preferences should be more broadly redefined to incorporate individual psychological determinants of betrayal reactivity and perceptions of social closeness.

Session 6: Trust | Room 2

Wednesday July 3, 2024 09:30 - 09:45

**The Impact of Reputation on Trust: Examining Individual Differences in Using Third-Party Evaluations in Cooperation Contexts**

Mathias Twardawski | Lucas John Emmanuel Köhler | Mario Gollwitzer

Department of Psychology, Ludwig-Maximilians-Universität München, Germany | Department of Psychology, Ludwig-Maximilians-Universität München, Germany | Department of Psychology, Ludwig-Maximilians-Universität München, Germany

People use different information to evaluate whether to trust somebody or not. A crucial construct in this regard is reputation: People form beliefs about someone's (un)trustworthy characteristics based on third-party evaluations. The present research investigates whether and how individuals use reputation information when deciding to trust potential cooperation partners and whether individual differences moderate this relationship. Specifically, we examine the roles of Victim Sensitivity—a dispositional sensitivity to environmental cues alluding to untrustworthiness and injustice to one's own disadvantage—and General Trust—the predisposition to trust others in social interactions regardless of specific contextual information. Across three pre-registered experiments (total N = 764), we explore how individuals differ in considering reputation cues for trust and cooperation intentions. We predict that people are more likely to trust and cooperate with others the better their reputation. This relationship should be moderated by people's Victim Sensitivity: Individuals high in Victim Sensitivity should only trust and cooperate with others when they have a high reputation. Conversely, General Trust should be positively related to trust and cooperation intentions irrespective of the reputation of the potential cooperation partner. To test these hypotheses, we presented participants with the decision to rent out their car to others for short-term use in exchange for money on a fictional platform. They received four requests from profiles of potential users, each rated on a 5-star scale by other platform users. We experimentally manipulated the reputation of the potential cooperation partner, varying from being new to the platform (no ratings), having a low reputation (3/5 stars), a medium reputation (4/5 stars), to a high reputation (5/5 stars). For each request, participants indicated their level of trust and the extent to which they would rent out their car to this person. Our findings show that General Trust predicts participants' decision to trust (or distrust) potential renters independent of the renters' reputation. By contrast, Victim Sensitivity predicts the extent to which participants use reputation information: individuals high in Victim Sensitivity are more sensitive to reputation information than individuals low in Victim Sensitivity. These insights illuminate the nuanced interplay between stable personality traits and the evaluation of reputation cues in cooperative decision-making.

Session 6: Trust | Room 2

Wednesday July 3, 2024 09:45 - 10:00

**They want and they can: Do decisions in shared social dilemmas depend not only on expectations of others' goodwill but also on others' competence?**

Stephan Nuding | Mario Gollwitzer

LMU, Munich | LMU, Munich

Trust can be defined as the willingness to accept a risk and become vulnerable based on the expectation of others' positive intentions or behaviour. Former research on trust in social dilemmas focused on the expectation of others' positive intentions and goodwill. Yet, expectations of others' behaviour could also be influenced by others' ability and competence to cooperate. In two preregistered studies (n1 = 302, n2 = 559), we tested whether the expectation of others' competence predicts cooperation in a social dilemma over and above the expectation of others' intentions. Participants played an online private-collective goods game with fictitious (Study 1) and real (Study 2) other participants. Participants had to reach a pre-specified threshold (otherwise, they would lose all their resource units) – either by investing their resource units into a private solution or a collective solution. The collective solution requires trust into the goodwill of the other players (i.e., the absence of free-riding), but potentially also trust into their (mathematical and strategic) competencies. Expectations of others' intentions and competence were measured via self-report scales in both studies and experimentally manipulated between-subject in Study 2. As hypothesized, more positive expectations of others' intentions were associated with higher contributions to the collective solution and lower contributions to the private solution in both studies. More positive expectations of others' competence were associated with lower contributions to the private solution in Study 2; yet, competence expectations were unrelated to contributions to the collective solution in both studies. Unexpectedly however, neither the experimental manipulation of intentions nor the experimental manipulation of competence in Study 2 had an effect on participants' contributions (neither to the private nor the collective solution). Yet, the experimental manipulations did have the expected effects on perceived expectations of intentions and competence. The findings are discussed with regard to potential avenues for further research.

Session 6: Trust | Room 2

Wednesday July 3, 2024 10:00 - 10:15

**The prevalence and magnitude of the correlation between generosity and trustworthiness: a meta-analysis of economic game experiments**

Wojtek Przepiorka | Ruohuang Jiao | Nas Momeni

Utrecht University, Department of Sociology | Utrecht University, Department of Sociology | Utrecht University, Department of Sociology

Rationale: Trust has long been recognized as one of the most important ingredients of social and economic life. People's expectation of others' trustworthiness is the precursor of these people's trust in others. The answer to how people form trustworthiness expectations is therefore key to answering why they (do not) trust others in particular situations. Theoretically, people's trustworthiness can be inferred from any behavioral or contextual signs the occurrence of which is (1) correlated with these people's trustworthiness and (2) regarded as indicative of trustworthiness by observers. Experimental research shows that generosity could be such a sign because it is correlated with trustworthiness and used to infer trustworthiness in social exchange with trust at stake. However, little is known about the prevalence and magnitude of the correlation between generosity and trustworthiness across different study designs and locations.

Methods: We conduct a meta-analysis based on individual-level data from 36 studies in which generosity and trustworthiness were measured within subject (individual-level N = 9765). In these studies, generosity was measured by means of donations to charity, dictator game giving and first-mover transfers in ultimatum games; trustworthiness was measured by means of second-mover decisions in (binary) trust games, sequential prisoner's dilemmas and second-mover transfers in investment games. We compare correlations between generosity and trustworthiness produced in two types of situations: (1) strategic situations, in which subjects know that their generosity could affect their interaction partners' decisions and in turn affect their payoffs (e.g., ultimatum game) and (2) "natural" situations, in which subjects' generosity cannot have such an effect by design (e.g. dictator game).

Results and conclusions: Overall, generosity and trustworthiness are moderately correlated ($r = 0.35$, $p < 0.001$). The correlation coefficients range from 0.02 to 0.69. Based on meta-regression analysis, we find that the correlation between strategic generosity and trustworthiness is smaller than the correlation between natural generosity and trustworthiness (coef. = 0.087, $p = 0.002$). These results substantiate that generosity can serve as a sign of trustworthiness and even more when generosity is exhibited naturally, i.e. without potential future rewards for generous acts. Our findings suggest that providing people with more opportunities to engage in and observe acts of generosity may promote trust in their neighborhoods and society at large.

Session 6: Trust | Room 2

Wednesday July 3, 2024 10:15 - 10:30

**Do people have a “trust propensity”? On the relation between common elicitation methods of trust**

Eyal Ert | Tamar Kugler | Eliran Halali | Nir Milstein

The Hebrew University of Jerusalem | University of Arizona | Bar Ilan University | Bar Ilan University

Rationale: Trust facilitates successful human interactions, with studies assuming the existence of a "trust propensity" influenced by specific contexts. Many different trust measures exist, from self-reported surveys to strategic games, yet their interrelationships remain largely unexplored. Previous investigations, mainly examining the relation between the GSS/IVS trust question and trust games, yielded mixed results. This study seeks to directly scrutinize the relationship between common trust measures and the overarching concept of a stable "trust propensity." We examine the possibility of an overarching trust factor ("T") could capture shared elements across measures. Alternatively, if measures were entirely independent, it would challenge the conventional notion of a universal "trust propensity," suggesting only context-specific influences on trust.

Method: Online pre-registered experiment (https://aspredicted.org/see_one.php), where 137 students completed a battery of trust elicitation tasks. Behavioral measures included: Trust game (Berg et al., 1995), Binary trust game (Fetchenhauer & Dunning, 2009), and self-reported past-trusting behavior (Glaeser et al., 2000). Attitudinal measures included GSS/IVS trust question, SOEP trust scale (Naef & Schupp, 2009), Trust in management scale (Mayer & Davis, 1999), and interpersonal trust scale (Dunn et al., 2012). Participants first played the trust game as player 1, then as player 2 using the strategy method. They were randomly matched for payment (half as player1, half as player 2). To assess temporal stability, the study was run three times with the same participants and tasks, in intervals of about 4 weeks apart.

Results: A cross-task analysis showed that the original and binary-trust games were highly correlated, though behavioral measures were uncorrelated with the attitudinal measures. The GSS/IVS and SOEP trust measures were significantly correlated with both types of measures (correlations ranged between 0.17- 0.21). Mean correlation within tasks across waves was 0.617, suggesting high within-task consistency. A factor analysis of the total correlation matrix confirmed a single factor across the different measures, with the general trust attitudes (GSS/IVS and SOEP) having the highest loadings (0.7). A temporal stability analysis of all 3-waves sessions revealed significant correlations for all trust measures over time (mean within-task-across-waves correlation range of 0.43 - 0.75), suggesting high temporal consistency.

Conclusions: The results suggest that behavioral trust measures are correlated, and that behavioral trust seem somewhat different than attitudinal trust. An interesting exception is the general question of trust (GSS/IVS) that correlates with all measures. Together, the result suggests preliminary evidence for “trust propensity” while alluding to the role of specific contexts.

Session 7: Cross-cultural | Room 1

Wednesday July 3, 2024 11:00 - 11:15

**Institutional trust and closeness-based favoritism across 25 societies**

Giuliana Spadaro | Vanessa Clemens | Andreas Glöckner

Vrije Universiteit Amsterdam | University of Cologne | University of Cologne

Individuals around the world display a tendency to favor members of their own groups compared to other groups when making decisions to trust and cooperate with others. According to recent theorizing, high-quality institutions can reduce ingroup favoritism by creating a safe context that reduces the risk associated with cooperating with others outside of one's own immediate entrusted network. However, while some studies find evidence in line with this idea, recent cross-national investigations find mixed-to-no support for the link between institutional quality and favoritism toward national ingroup members. We argue that this might be due to how favoritism was assessed in previous studies (national group membership-based) and the fact that state-level institutions might not be relevant for cross-national exchanges. In this project, we analyzed within-country interactions in which we assessed favoritism based on social closeness. Accordingly, we tested whether high institutional trust is associated with lower closeness-based favoritism in trust and cooperation.

We tested these hypotheses across two preregistered studies conducted in the US (N = 336) and across 25 societies around the globe (N = 6,182). To assess closeness-based favoritism in trust and cooperation, we asked participants to interact with a person they consider themselves close to (i.e., close other) and with a stranger from their own nation (i.e., national stranger) in two hypothetical one-shot continuous prisoner's dilemma games (PD). Transfer in the PD was used as a measure of cooperation, expectations of the other person's transfer were used as a measure of trust. To assess perceived institutional quality, we assessed individuals' trust in their national institutions. Mixed effect models were used across all analyses to account for the nested structure of the data.

In both studies, we find evidence for closeness-based favoritism in trust and cooperation, indicated by significantly higher cooperation and trust towards close others compared to national strangers. In line with our preregistered hypotheses, we find a significant negative interaction between interaction partner and institutional trust for both trust and cooperation. This suggests that individuals with higher institutional trust show lower closeness-based favoritism in both trust and cooperation.

In two studies, we provide converging evidence that support the role of institutions in shaping trust and cooperative behavior. Consistently across 25 societies, higher institutional trust was associated with lower closeness-based favoritism. Our findings also underline the importance of taking into account individuals' perceptions (in terms of interpersonal closeness and institutional trust) as potential drivers underlying favoring behavior.

Session 7: Cross-cultural | Room 1

Wednesday July 3, 2024 11:15 - 11:30

**The impact of loneliness on economic trust: experimental evidence from 27 European countries**

Elena Stepanova | Marius Alt | Astrid Hopfensitz

Joint Research Centre European Commission | Joint Research Centre European Commission | emlyon business school

Trust behavior and being trusted are influenced by a multitude of individual and situational factors. In this paper, we focus on a novel dimension, hypothesized to be related to trust behavior, that has so far received little attention in economics: loneliness.

Through a large, incentivized trust experiment conducted in 27 European countries with more than 27000 respondents, we investigate: (i) the relationship between self-reported loneliness and trust and trustworthiness behavior and (ii) the impact of loneliness on receiving trust from others.

In line with previous research from psychology, we observe a strong negative correlation between self-reported general trust and loneliness. This relationship is however not replicated in an incentivized trust setting: lonely individuals are even more trusting than individuals who are not lonely. Lonely individuals seem no different from non-lonely individuals regarding their trustworthiness. We finally observe that lonely individuals are treated significantly differently in the trust game: they receive significantly more trust from others and benefit from more trustworthy behavior.

Overall, our results suggest that lonely individuals are willing to trust more than non-lonely individuals when real monetary stakes are at hand. This increased willingness to take social risk is however not reflected in their own self-reports regarding trust.

Session 7: Cross-cultural | Room 1

Wednesday July 3, 2024 11:30 - 11:45

**Social Control across 60 Societies: self-interested norm enforcement versus strong reciprocity**

Kristen Syme | Alice Saraiva Angra De Oliveira | Daniel Balliet

Institute of Security and Global Affairs, Leiden University, the Hague, the Netherlands | Department of Social and Organizational Psychology, VU University Amsterdam, the Netherlands | Department of Social and Organizational Psychology, VU University Amst

This preregistered study tests predictions derived from two competing evolutionary theories of second- versus third-party punishment using a sample of 800 ethnographic texts from 60 societies in the Probability Sample Files (PSF) of the Human Relations Area Files (HRAF). Strong reciprocity (SR) is a theory of human cooperation that proposes that cooperative norms are maintained, in party, by a willingness of third-party norm enforcers to punish unfair behavior regardless of the immediate or future benefits to themselves (Fehr, Fischbacher, and Gächter, 2002). An alternative theory, the self-interested enforcement hypothesis (SIE), proposes that cooperative norms are maintained by individuals pursuing their self-interests (Singh & Glowacki, 2017). In line with SIE and contra SRE, we predicted that third-party enforcement would be associated with indices of power based on group size, status, sex, and age. Texts were extracted from the HRAF database using OCM code 626 on 'social control'. The dataset contains over 2,000 sections of text on social control from all societies in the PSF. Texts were randomly assorted, and the first 800 sections were coded by two coders who reached consensus on the coded variables. One of the coders was blind to the hypotheses. The coded variables measure enforcers, offenders, and harmed parties on dimensions of power. Using inductive methods, we also coded enforcement types (e.g., gossip, ostracism) and offense types (e.g., assault, sexual infidelity). Descriptive results show a greater frequency of second-party enforcement (33.9%) compared to third-party enforcement (22.1%); the remaining text accounts were coded as both or unknown. As predicted, third-parties tend to score higher on indices of power compared to second-parties. A greater percentage of third-party enforcers compared to second-party enforcers are groups (66.5%, 47.4%), status holders (77.8%, 27.8%), males (36.9%, 29.6%), and adults (97.8%, 90.2%), whereas among offenders, 16.4% are groups, 12.0% are status holders, 28.5% are male, and 86.4% are adults. These preliminary findings lend support to the hypothesis consistent with SIE that enforcement is associated with self-interest and power. I will discuss these and the results of the preregistered analyses in my talk.

Session 7: Cross-cultural | Room 1

Wednesday July 3, 2024 11:45 - 12:00

**Corrupt Collaboration in 20 Societies**

Jonas Ludwig | Jonathan Schulz | Shaul Shalvi | Ivan Soraperra | Ori Weisel

Technische Universität Berlin | George Mason University | University of Amsterdam | Max Planck Institute for Human Development | Tel Aviv University

Rationale: Humans are an especially cooperative species (Bowles & Gintis, 2011). We are also remarkably honest, even when tempted to lie to secure personal profit (Abeler et al., 2019; Gerlach et al., 2019). A recent body of work suggests that when honesty and cooperation are at odds (Weisel & Shalvi, 2015), honesty often gives way to corrupt collaboration. In the current project we explored how the balance between honesty and cooperativeness depends on the cultural setting, connecting to a growing body of research demonstrating significant cultural variation in ethical thinking and behavior (Gächter & Schulz, 2016; Schulz et al., 2019). Since corrupt collaboration requires both dishonesty and cooperation, which generally correlate negatively, it is difficult to predict how it will vary across societies.

Method: We conducted an online study in 20 countries (N=11,500). The study included measures of individual (dis)honesty (using a variation of a die-rolling task), cooperativeness (using a public good game), dyadic dishonesty (using a variant of the dyadic dishonesty task, Leib et al., 2021), and a host of additional individual measures (cognitive reflection task, honesty-humility scale, trust, family ties, tightness, socioeconomic stratification, categorization type, individualism-collectivism).

Results: We are at the initial stage of exploring the rich dataset. Initial analyses reveal considerable variation in collaborative dishonesty across 20 societies; individual dishonesty is positively correlated with collaborative dishonesty, while individual cooperativeness is negatively related to collaborative dishonesty; kinship intensity interacts with both individual dishonesty and cooperativeness to predict collaborative dishonesty.

Conclusions: Similar to other behaviors, e.g., individual dishonesty and cooperativeness, the tendency to engage in corrupt collaboration varies between societies. It is positively related to individual dishonesty – people who tend to lie for personal profit also tend to engage in dyadic dishonesty; and negatively related to cooperativeness. The latter result is of particular interest: Although a successful, mutually beneficial relationship with a partner, which is based on joint dishonesty, clearly requires cooperation, a cooperative disposition, which is related to a more general tendency toward moral behavior, seems to hinder such joint, corrupt enterprises.

Session 7: Cross-cultural | Room 1

Wednesday July 3, 2024 12:00 - 12:15

**Guilt- and Shame-Driven Prosociality Across Societies**

Catherine Molho | Ivan Soraperra | Jonathan F. Schulz | Shaul Shalvi

VU Amsterdam | Max Planck Institute for Human Development | George Mason University | University of Amsterdam

Impersonal prosociality is considered a cornerstone of a thriving civic society and well-functioning institutions. Previous research has documented substantial cross-societal variation in impersonal prosociality. Such variation might arise from cultural evolutionary processes resulting in different moral systems—that is, packages of psychological mechanisms, norms, and institutions that regulate social behavior. Here, we test the idea that different mechanisms support prosociality in societies with diverging moral systems. Specifically, we propose that, in some societies, prosociality is based on guilt and internalized moral norms whereas, in other societies, it is based on shame and external enforcement. Based on this rationale, we hypothesize that, in guilt-prone societies, prosociality is more strongly affected by providing information about the (negative) consequences of one's decisions on others. In contrast, in shame-prone societies, prosociality is more strongly affected by making decisions observable.

In a registered report, we proposed a cross-societal experiment in 20 culturally diverse countries around the world. In each country, we will recruit a sample size of 390 participants, for a total of 7,800 participants across countries. Previous cross-societal research on prosociality relied on economic decision-making tasks, where (1) individuals could allocate money between themselves and others, (2) received full information about how their decisions impact others, and (3) made their decisions privately. To examine how guilt and shame affect prosociality across societies, we will administer dictator games and introduce two variations that allow us to tease apart the effects of guilt and shame on prosocial decisions. First, to experimentally induce guilt, we will use a 'wilful ignorance' paradigm, and vary information about the consequences of one's decisions on others (full versus hidden). Second, to experimentally induce shame, we will vary the observability of decisions (public versus private). We will also measure guilt- and shame-proneness at the individual and societal level.

Across societies, we expect more prosocial choices when participants make decisions under full rather than hidden information, and when they make decisions publicly rather than privately. Importantly, we also hypothesize that the effects of information and observability will vary across individuals and societies. We expect that activating guilt by varying information more strongly increases prosociality among guilt-prone individuals and societies. In contrast, we expect that activating shame by varying observability more strongly increases prosociality among shame-prone individuals and societies. Our study provides a first comprehensive investigation of how guilt versus shame affects prosociality across societies and sheds light on cultural universality and diversity across understudied populations.

Session 7: Cross-cultural | Room 1

Wednesday July 3, 2024 12:15 - 12:30

**Antecedents and consequences of cross-national social preferences**

Vanessa Clemens | Marina Orifici | Angela Dorrough | Laura Fröhlich | Andreas Glöckner

University of Cologne | University of Cologne | University of Cologne | FernUniversität Hagen | University of Cologne

Study rationale: To combat current global crises such as climate change, cross-national cooperation – i.e., cooperation between individuals from various nations – is crucial. Core determinants of cooperation are an individual's social preferences. Previous work on social preferences in cross-national studies has shown that they vary according to both, the individual's and their interaction partner's nationality (i.e., cross-national social preferences). Building on previous work, we systematically investigate further characteristics of the individual, the interaction partner, and the specific nation-dyadic combinations associated with variation in cross-national social preferences. We tested hypotheses concerning the role of ingroup membership, cultural similarity, perceived stereotypical similarity, the economic status of the interaction partner's nation (in relation to one's own), and (militarized) conflicts between the nations as antecedents for cross-national social preferences. Furthermore, we test whether an individual's cross-national social preferences predict cross-national cooperation.

Methods: We conducted a preregistered, fully-incentivized multi-national study (N = 6,182; 154,550 decisions) across 25 nations. Participants were asked to indicate their cross-national social preferences towards individuals from the participating 25 nations using an adapted version of the standard social value orientation slider. We then presented participants with prisoner's dilemma games to assess cross-national cooperation with individuals from the respective 25 nations. Cultural similarity, economic development, and militarized conflicts were assessed using indicators on a national level. Nation-specific stereotypes were assessed on an individual level. For our analysis, we applied linear mixed models to account for the nested structure in our data.

Results: We find evidence for national ingroup favoritism, indicated by significantly higher cross-national social preferences for individuals of one's own compared to other nations. In line with our hypotheses, we show that individuals have significantly higher cross-national social preferences for those that are culturally and stereotypically similar. Individuals from wealthier nations display higher cross-national social preferences towards individuals from less wealthy nations. Cross-national social preferences decrease as the number of militarized conflicts between the respective nations increases. The results remain robust when controlling for an individual's general social preferences. Importantly, as hypothesized, we show that cross-national social preferences predict cross-national cooperation even when controlling for expectations of cooperation.

Conclusion: We report evidence for cross-national social preferences that vary according to characteristics associated with the specific dyadic nation combinations (e.g., similarity, militarized conflicts). Importantly, we show that these cross-national social preferences predict

cross-national cooperation. The findings underline the importance of investigating specific nation-dyads to better understand what drives cross-national cooperation.

Session 8: Conflict & Competition | Room 2

Wednesday July 3, 2024 11:00 - 11:15

**Unveiling the hidden impact of the conflict structure on parties' settlements. A field-level experimental black-box approach from 1971 to 2021**

Johann M. Majer | Martin Schweinsberg | Marco Warsitzka | Hong Zhang | Roman Trötschel

University of Hildesheim | ESMT Berlin | Leuphana University | Leuphana University

The field of conflict resolution research seeks to understand, mitigate, and resolve conflict. However, the field overlooked a fundamental situational feature affecting parties' settlements: The (objective) conflict structure. We don't even know whether high or low levels of conflict facilitate conflict settlement. Drawing on five decades of research (1971-2021), we analyze the field's experimental findings to identify how the conflict structure affects conflict resolution. We examine data from N=30,361 conflict simulations and N=52,025 counterparts (from N=234 primary studies). The results neither show a linear nor a nonlinear impact of the conflict structure on settlements. Instead, our analysis reveals that sharing compatible interests moderates the impact of the conflict structure on settlements. When parties do not share compatible interests, increased conflict improves settlements. However, when parties do share compatible interests, increased conflict hinders settlements. Distinct psychological routes towards settlements via increased motivation or impaired cognition likely account for the observed moderation effect. Our field-level experimental black-box approach identifies the precise conditions of the conflict structure that hinder and enhance settlements. These findings integrate the two major but isolated subfields of cognition and motivation by revealing when and how they distinctly affect settlements. We show the objective conflict structure is a highly generalizable condition that affects conflict resolution. The conflict structure can bridge theory and practice by helping scholars better understand under which conditions of conflict what behavioral intervention can help resolve conflict and why.

Session 8: Conflict & Competition | Room 2

Wednesday July 3, 2024 11:15 - 11:30

**Interpersonal conflict resolution crowds out forward-looking decision making**

Johann M. Majer | Laura E. M. Stalenhoef | Hillie Aaldering | Caroline Heydenbluth | Roman Trötschel

University of Hildesheim | University of Hildesheim | Free University of Amsterdam | Leuphana University | Leuphana University

Addressing societal challenges requires interpersonal conflict resolution with opponents and forward-looking consideration of future interests. These two central psychological challenges have been studied in isolated research silos. However, we posit that interpersonal and intrapersonal conflicts are psychologically more similar than previously suggested and their solutions are interdependent. Parties have to resolve this interplay of conflicts at the same time. Building on the fields of conflict resolution and decision making, we hypothesize that parties prioritize conflict resolution with the present counterpart leading parties to sacrifice own future interests. Across five simulated and interactive experimental conflict studies, online and face-to-face, with and without incentives, we find empirical support for our prioritization hypothesis. The simulated studies 1a and 1b suggest that conflict parties tend to make proposals that resolve competing interests in a present interpersonal conflict over intrapersonal conflict and interests in an intrapersonal conflict over interpersonal conflict in the future. Study 2 provides support for parties' priority on the present interpersonal conflict when three interdependent conflicts emerge at the same time. The interactive experiments 3 and 4 provide further support for prioritizing conflict resolution with the present counterpart over the consideration of future interests. Contrary to previous findings, the prioritization effect becomes stronger over the course of conflict resolution. Our research extends current theorizing by integrating previously isolated conflict resolution and decision-making research. We discuss the psychological roots of the observed prioritization effect and how conflict parties can master the real-world challenge of conflict resolution while keeping an eye in the future.

Session 8: Conflict & Competition | Room 2

Wednesday July 3, 2024 11:30 - 11:45

**Testosterone affects learning of implicit social dominance hierarchies through competitive interactions**

Annabel Losecaat Vermeer | Romain Ligneul | Gabriele Bellucci | Rémi Janet | Soyoung Park | Jean-Claude Dreher | Claus Lamm

Neuropsychopharmacology and Biopsychology Unit, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Vienna, Austria; Social, Economic and Organisational Psychology, Institute of Psychology, Leiden

Rationale: Learning your position and that of others within a social dominance hierarchy is vital for successful social interactions. Androgen testosterone has been associated with social dominance behaviour, monitoring social threats, and reduced punishment sensitivity. Moreover, testosterone has receptors in brain regions associated with reward and learning of dominance hierarchies such as medial prefrontal cortex (mPFC), amygdala, hippocampus, and striatum. Here, we examined how exogenous testosterone influences the learning of dynamic social dominance hierarchies and competitive decision-making.

Methods: To test this, forty-five males (M = 25.2 years, SD = 3.64) partook in a double-blind, placebo-controlled, crossover fMRI study. Participants learnt an implicit dominance hierarchy while playing a competitive task against three opponents of different skill. Each round, participants had to decide whom they wanted to compete against and received feedback on whether they won or lost. As a control, participants played a non-social reinforcement learning task. To assess learning, we used a selection of reinforcement-learning models, and a softmax decision rule to translate opponent-specific dominance values into choice probabilities.

Results: Results revealed that participants learnt the dominance ranks as shown by an increased choice preference for the lower ranked (i.e. weaker) opponent of each pair ($p < .001$). Compared with placebo, testosterone did not influence choice behaviour for specific opponents but affected the learning of their dominance ranks as shown by model comparison ($p < .05$). A model including different learning rates for wins and losses in updating the dominance value of the specific opponent showed that exogenous testosterone (vs. placebo) decreased the learning rate for losses over wins ($p < .01$). Importantly, testosterone (vs. placebo) did not differentially influence learning during the non-social task ($p < .05$). Analysis of brain data revealed increased activity of the ventral mPFC, striatum and amygdala for social wins over social losses. The opposite contrast exhibited increased activity of the dorsal mPFC and anterior insula. Competitive prediction errors (PE) were tracked by the mPFC similarly in both treatment groups.

Conclusion: To conclude, testosterone caused asymmetric updating of the dominance ranks of social opponents following wins and losses. Testosterone did not affect learning in the non-social task. This suggests that testosterone specifically reduces the sensitivity to losses in competitive interactions when establishing one's own social dominance. Brain responses to social outcomes and tracking of competitive PEs were not modulated by testosterone.

Session 8: Conflict & Competition | Room 2

Wednesday July 3, 2024 11:45 - 12:00

**The Cultural Logic of Honor, Competition and Coordination**

Shuxian Jin | Ayse K. Uskul | Angelo Romano | Vivian L. Vignoles | Alexander Kirchner-Häusler | the HONORLOGIC Team

University of Sussex | University of Sussex | Leiden University | University of Sussex | University of Sussex

Human societies face many challenges that require collective action for resolution. Researchers has used economic games to investigate how people interact with others to achieve collectively beneficial outcomes at a personal cost within and across cultural contexts. However, little is known about how individuals from cultures that promote honor as a core value manage conflicts between self-interests and concern for others. As a unique form of motivational systems, honor contains various components such as self-promotion, retaliation, and concerns for family reputation, and may drive behaviors that are clustered around pursuing the self-worth of honor. The present study investigates whether honor can drive 1) competitive responses of individuals in situations where one can be better off at the cost of the other, and 2) withholding choices in situations where one may risk wasting effort in coordinating with others for better outcomes. We tested pre-registered hypotheses using an online experiment where participants (from 13 societies; N = 3,371) made incentivized competition decisions in the contest game and coordination decisions in the step-level public goods game. Honor was measured as personal endorsement and perceived societal-prevalence of honor values. Data were collected from the Mediterranean region, including Spain, Italy, Greece, Turkey, Cyprus, Lebanon, Egypt, and Morocco, as cultural contexts where honor has been shown to be a value deeply ingrained in individuals' social worlds (albeit in different degrees), and from East Asian (Japan and Korea) and Anglo-Western societies (the US, the UK) to provide a broader comparative perspective. Results indicated that societies where honor values are perceived to be more prevalent are associated with more competition and more coordination in social interactions with strangers. At the individual level, honor was separated into two dimensions, including the defense of family reputation and the self-promotion and retaliation. The perceived-societal prevalence of both dimensions positively relate to competition and coordination, but personal endorsement of honor values showed mixed relations with these behaviors. Our findings highlight the coexistence of positive relation between the perceived-societal prevalence of honor and both competitive and coordinative behaviors. This suggests that honor may function in dual capacities, positively and negatively influencing the resolution of challenges that demand collective efforts. The discrepancy in findings regarding the personal endorsement of honor values raises question about the underlying mechanisms through which different components of honor may drive behaviors. This study enhances our understanding of the role of honor in shaping important interpersonal and societal processes.

Session 8: Conflict & Competition | Room 2

Wednesday July 3, 2024 12:00 - 12:15

**Expected outcomes and risk-taking in competitive contexts: A large-scale analysis of gambits in tournament chess**

Uri Zak | Allon Vishkin | Eldad Yechiam

The Hebrew University of Jerusalem | Technion – Israel Institute of Technology | Technion – Israel Institute of Technology

Life is rife with competition, and competition is rife with risk-taking decisions. When individuals compete with each other for academic grades, job promotions, sports achievements, status, or popularity – what drives their risk-taking? One factor of particular interest is competitors' expected outcomes. If expecting better (vs. worse) outcomes is parallel to the domain of gains (vs. losses) in a lottery, it could be related to reduced risk-taking (e.g., Kahneman & Tversky, 1979). On the contrary, if expecting better (vs. worse) outcomes coincides with expecting risks to pay off, it could be related to enhanced risk-taking. To investigate this issue, we studied how expected outcomes are related to risk-taking in the context of tournament chess.

Millions of people around the world play tournament chess and many of them are rated by the international chess rating system. The rating difference between two players predicts the average outcome (with a larger difference predicting a more one-sided average outcome), and players' ratings continuously change according to their actual results to maintain the reliability of predictions based on ratings. Ratings are also salient: they are posted on billboards in playing halls and are readily available online. The reliability and accessibility of ratings make tournament chess an ideal setting to study initial expected outcomes. To assess risk-taking, we utilized the concept of a gambit, a chess opening in which a player gives up a piece (or more) in favor of a positional edge. The asymmetry between a material loss and a positional gain in gambits is considered risky by the chess community at all levels of play. Reasonably, it leads to a game where both sides have more winning opportunities (and a draw is less likely). We specifically validated this claim in our dataset, which included 5,676,486 games played by 225,200 players.

We found robust evidence that the relationship between expected outcomes and risk-taking is positive. Larger rating differences between opponents were associated with a higher probability of risk-taking by playing a gambit. This effect was sizeable: For example, increasing the rating difference by 200 points (roughly one SD) was estimated to enhance risk-taking by approximately 9.2%. A similar effect was observed when adding a variety of control variables and using a within-player analysis. To conclude, competitors tended to take more risks when expecting better outcomes.

Session 8: Conflict & Competition | Room 2

Wednesday July 3, 2024 12:15 - 12:30

## On the Context-Specificity of Ideological Asymmetries in (Intergeroup) Mistrust and Aggression

Ruthie Pliskin | Laura Hoenig | Jörg Gross | Carsten de Dreu

Leiden University | Leiden University | University of Zurich | Leiden University and the University of Amsterdam

Previous research on ideological differences in mistrust and aggression has often arguing that rightists are more inclined than leftists to both—especially towards outgroup members—but these claims have been countered by research in which such differences do not emerge. Importantly, past research has never fully isolated relevant political cues from the examination of such ideological differences, nor compare ideologically-neutral situations to politically-laden ones. In two studies, we employed experimental game (the Attacker Defender Contest, measuring mistrust; and the Joy of Destruction Game, measuring aggression) to overcome these gaps. UK and US-based Participants (Study 1 n = 399; Study 2 n = 769) made fully-incentivized monetary decisions opposite strangers (i.e., neutral condition), but also opposite counterparts identified in terms of their ideology or their nationality (i.e., ingroup and outgroup conditions). No ideological differences emerged when politically-relevant cues were absent. When facing ideologically-identified targets, however, leftists showed more parochial behavior than rightists, differentiating ingroup and outgroup members to a greater extent. The type of ingroup and outgroup also mattered, such that when categorization was nationality-based, rightists were slightly more parochial, mistrusting and aggressive than leftists. We discuss implications for the debate on and study of ideological differences and similarities.

Session 9: Prosociality in the wild | Room 1

Wednesday July 3, 2024 14:00 - 14:15

**Preferences for Status Quo, Adaptation or Mitigation in Collective Risks Situations**

Federica Maria Raiti | Gianluca Grimalda | Nancy Buchan

Sapienza University of Rome | University of Passau | University of South Carolina

Rationale: Collective risk social dilemmas are interactions where individual contributions lead to decreased probability that collective losses will occur, as is the case with climate change. In spite of their importance, our knowledge of how individuals interact in such interactions is still scant.

Methods: Our research took place in rural areas of Papua New Guinea, where people are highly exposed to the risks of climate change. In the Game condition, pairs of participants choose between three lotteries in which (a) they face a high monetary loss with an 80% probability, or they spend some of their endowment to (b) reduce the monetary loss, or (c) reduce the probability of the negative event occurring. If an individual opts for lottery (c), the probability of losses is also reduced for the other individual. We interpret the choice of (b) - or (c) - as preference for adaptation to – or mitigation of - collective risks. In the Individual condition, individuals are faced with a choice with the same lotteries as in the Game condition, the only difference being that the choice of (c) does not affect the probability of loss of another player.

Self-interested individuals should be indifferent between the two conditions. Altruistic individual (or conditionally cooperative individuals) should choose lottery (c) with higher frequency in the Game than the Individual condition in all cases (or if they expect their counterpart to choose lottery (c) with a sufficiently high probability. Betrayal-averse individuals should do the opposite.

We construct lottery (b) and (c) to have the same expected value with lottery (c) having higher variance. Lottery (a) has lower expected value than (b) and (c).

Participants in the Game condition played the game twice with an ingroup (a person from the same village) and an outgroup.

Results: Individuals choose lottery (c) more often in the Game than in the Individual condition. Lottery (b) is the modal choice, but about 30% of the sample choose lottery (a). No ingroup effect is observed.

Conclusions: A significant share of individuals construe the collective risk social dilemma as a cooperative choice. The fact that a sizable portion of the sample chooses lottery (a), presumably because it yields the highest payoff in case of “win”, suggests a tendency of individuals not to try any adaptation nor mitigation even under the threat of severe losses.

Session 9: Prosociality in the wild | Room 1

Wednesday July 3, 2024 14:15 - 14:30

**Market integration, egalitarianism, and reward of merit: An experimental analysis from Papua New Guinea small-scale societies**

Gianluca Grimalda | Matthias Sutter | Andreas Pondorfer

Passau University | Max Planck Institute for Collective Goods | Technical University of Munich

Rationale: Some literature posits that market interactions help develop a sense of generalized pro-sociality, which extends to complete strangers the sense of particularized pro-sociality normally reserved to one's kins. Another approach posits the opposing view that market interactions erode altruism.

Methods: We sampled 1800 participants from 30 villages of Bougainville, Papua New Guinea. Communities are transitioning from self-subsistence to market integration, thus providing a suitable setting to test the above hypotheses. We measure prosociality through the "power-to-take" game. Two stakeholders are assigned different individual earnings based on their relative performance in a tournament (Merit condition) or luck in a random draw (Luck condition). They both make a proposal on how to split the total earnings. A "spectator" also proposes a division between the two stakeholders. One of the three choices is randomly selected and implemented. We measure market integration through (a) the share of calories coming from market purchases vis-à-vis self-subsistence, and (b) the revenues from selling one's crops on international markets. Distance from and costs of travelling to the closest market town are ancillary measures of integration.

Results: We find a weak tendency for greater market integration to be associated with lower pro-sociality. In particular, the egalitarian division of the pie is widespread in remote communities lacking access to drivable roads. Selfish behaviour is more frequent in communities with fast road connection to the market town. No merit effect is observed.

Conclusions: The correlation between market integration and prosociality appears weak. Social norms supporting greater tolerance of inequality and acceptance of market transactions rather than traditional networks of reciprocal exchange mediate such a relationship.

Session 9: Prosociality in the wild | Room 1

Wednesday July 3, 2024 14:30 - 14:45

## Highlighting the Collective Harm: Tackle Illicit Drug Use in the Netherlands With Moral Appeals

Sylvia. Y. Xu | Laetitia. B. Mulder | Floor. A. Rink

University of Groningen | University of Groningen | University of Groningen

Addressing the pervasive issue of recreational illicit drug use remains a global challenge yet years of anti-drug campaigns yielded unsatisfactory results. This research aims to provide a novel intervention by framing recreational illicit drug use as a problem of social dilemma — a conflict between immediate self-interest (i.e., the pleasure of substance use) and longer-term collective interests, given its enduring detrimental impacts on societies (e.g., contributing to organized crime and cross-national corruption). While previous anti-drug campaigns have mostly focused on the adverse effects of illicit drugs individually (i.e., health-related harm), only a few have emphasized its collective harm. What are the effects of moral appeals highlighting the collective harm of illicit drug use on decreasing illicit drug use intention and increasing negative attitudes toward drug use? And how will prior illicit drug use moderate such effects? To answer these questions, we conducted four studies in the Netherlands with surveys and a longitudinal, field experiment.

Studies 1-3 (Ntotal = 917) tested the effect of moral appeals delivered in text format with online surveys. Study 4 (N = 613) employed a two-week time-lagged design, targeting participants attending events with prevalent recreational illicit drug use, and introduced video moral appeals alongside text format. Exposed to the manipulations in Wave 1, participants reported intentions and attitudes in both waves and disclosed actual drug use behaviors between Wave 1 and 2. Across studies, the results revealed the potential of moral appeals in reducing future illicit drug use intentions, particularly for individuals with a history of drug use. Additionally, the research revealed a broader impact of moral appeals in fostering negative attitudes toward illicit drug use, irrespective of individuals' prior behaviors. A subsequent mini meta-analysis further supported these findings.

In summary, by framing the use of illicit drug use as a social dilemma, our study illuminates the potential of moral appeals as a powerful intervention. By emphasizing collective harms, our findings offer a nuanced perspective and advocate for the incorporation of moral appeals in future anti-drug initiatives, marking a pivotal step toward more effective interventions.

Session 9: Prosociality in the wild | Room 1

Wednesday July 3, 2024 14:45 - 15:00

**Charitable Donation Theories in the Wild: Evidence from a Large Online Donation Platform**

Tom Gordon-Hecker | Coby Morvinski

Ben-Gurion University | Ben-Gurion University

As charitable giving grows, researchers delve into understanding why individuals give away their hard-earned income for the benefit of someone else. A specific feature of donation decisions that has piqued the interest of many researchers is the identifiability of the donation recipients. Here, we elaborate on previous findings using a rich dataset of over 5.5 million real donations made with respect to about one million fundraising campaigns. This dataset allows us to simultaneously test identifiability related theories, by exploring a highly ecologically relevant, yet understudied factors that might influence donation decisions, namely visual cues. Hence, this study explores if the visual identifiability of recipients, rather than their actual composition, impacts people's donation choices.

We use donation decisions made on a non-profit funding platform that allows individuals to donate to classroom project requests from public school teachers. We used a machine learning technique to detect and count the number of identifiable faces in the photos attached to the campaign. We tested three effects: the effect of including a photo (i.e., vividness), the effect of having any number of identified recipient/s in the photo (i.e., identifiability effect), and the effect of having a single identified recipient in the photo (i.e., singularity effect). The results revealed that: (i) Merely including a photo in the campaign increased success rate ($\beta = .371$, $p < .001$), as well as the number of donors ($\beta = .132$, $p < .001$), but not the average donation for the campaign ($\beta = .003$, $p = .282$). (ii) Having an identified recipient (one or more) in the campaign's photo increased the success rate ($\beta = .045$, $p < .001$), average donation ($\beta = .001$, $p < .001$) and number of donors ($\beta = .025$, $p < .001$), suggesting an identified victim effect. (iii) Featuring a single (vs. multiple) recipient in the campaign's photo negatively affected the success rate ($\beta = -.027$, $p = .005$) and the number of donors ($\beta = -.020$, $p = .005$), but not the average donation ($\beta = -.006$, $p = .128$), suggesting there was no singularity effect (if anything, the effect of singularity is negative).

Taken together, our results reveal that whereas some well documented findings from the lab (e.g., identified victim) are evident also in field data, others are not (e.g., singularity effect). This work also demonstrates how behavioral researchers can benefit from obtaining and analyzing large amounts of field data in order to validate and generalize existing behavioral theories. Beyond enhancing confidence in published theories, extending lab results to real-life contexts provides practitioners with crucial insights for practical applications in their domains.

Session 9: Prosociality in the wild | Room 1

Wednesday July 3, 2024 15:00 - 15:15

**Cooperation and Punishment in the general population: Evidence from a representative experiment in Germany.**

Silvia Angerer | Helena Baier | Daniela Glätzle-Rützler | Philipp Lergetporer | Thomas Rittmannsberger

UMIT Hall | Technical University of Munich | University of Innsbruck | Technical University of Munich | Technical University of Munich

Adhering to cooperative norms is pivotal in the functioning of societies, affecting everything from interpersonal relationships to global collaborations. A thorough understanding of the factors shaping norm-compliance is crucial for tackling collective challenges. In this study, we explore the effects of third-party punishment (TTP) on cooperative behavior, using a large (n=2,215) nationally representative sample of German adults aged 18-69 years. Unlike previous research, which often concentrates on specific groups and settings, using a sample that reflects the demographic composition of the population permits a nuanced examination of behaviors across subgroups while enhancing the external validity of the findings. Respondents played a one-shot prisoner's dilemma game, either with, or without the possibility of facing punishment for non-cooperation. Consistent with prior studies, we find a 6-percentage-point increase in cooperation rates when punishment is a possibility. Notably, respondents tended to overestimate the likelihood of punishment. While 59% of respondents believe in being punished for non-cooperation, only 22% were willing to engage in costly punishment. Our results revealed an association between age and behavior. Older respondents demonstrated significantly higher levels of cooperation compared to their younger counterparts, while the opposite is true for the decision to punish non-cooperators. Interestingly, factors such as gender, political affiliation, income, and migration background seem unrelated to behavior in the cooperation game, although residing in areas characterized by a low share of citizens with a migration background seems to increase the readiness to cooperate. Moving beyond the controlled environment of the experimental setting, we explored the connection between cooperation in the prisoner's dilemma game and real-world prosocial behavior. Our results suggest a positive link between cooperation in the experiment and engagement in honorary work, registration as an organ donor, and adherence to COVID-19 vaccination. These findings underscore the practical relevance of our results, demonstrating how behaviors observed in controlled experiments extend meaningfully to actions in everyday life. In conclusion, our study offers valuable insights into the effect of punishment on cooperation, revealing the positive impact of third-party punishment on prosocial behavior in the general public.

The study was pre-registered in the AEA RCT Registry (11936)

## Psychological Mechanisms Underpinning People's Willingness to Vaccinate Against Future Viruses

Hagai Rabinovitch | Tehila Kogut

University of Amsterdam | Ben-Gurion University of the Negev

The global COVID-19 pandemic has claimed over 4 million lives, affecting more than 190 million people to varying degrees. The vaccine has significantly reduced infections and saved millions of lives. Despite its proven efficacy and satisfactory safety profile, a notable percentage of opposers in many countries pose a risk to global health. Additionally, concerns arise about individuals feeling pressured to vaccinate due to external factors, potentially impacting future vaccination rates.

While the virus currently poses a reduced threat, the emergence of new subvariants or a vaccine-resistant virus is imminent. This study explores how thinking styles (intuitive and rational) influence perceptions of virus and vaccine risks, predicting COVID-19 vaccination decisions. It delves into the distinction between motivation stemming from free will and external pressure, a novel aspect in this context. Three groups are identified: unvaccinated, vaccinated under external pressure and vaccinated by free will.

Method: Two online studies (N = 652) were conducted, one during the pandemic, six months post-vaccine availability, and another after the subsiding of COVID-19 threats. Participants' vaccination status, willingness to vaccinate against future viruses, thinking styles and motivation for vaccination were measured.

Results: Unvaccinated individuals perceived vaccine risk as higher than virus risk, while vaccinated individuals exhibited the opposite pattern. A two-way interaction between vaccination status and thinking style reveals higher rational than intuitive scores among vaccinated individuals. Unvaccinated people utilize both thinking styles similarly. Intuitive thinking partially mediates the link between vaccination decisions and vaccine risk perceptions, suggesting that increasing rational thinking may reduce vaccine objection among the unvaccinated. Motivations to vaccinate vary; those choosing to vaccinate freely show higher rational scores, while those feeling compelled exhibit the reverse. Future vaccination intentions do not always align with past decisions. Individuals vaccinated freely are more likely to vaccinate again, whereas those feeling forced are less likely.

Conclusion: This research enhances the understanding of psychological mechanisms influencing vaccination behavior. Unvaccinated individuals, higher in intuitive thinking, may fear the vaccine more than the virus, impacting their behavior. Assumptions about continued vaccination among previously vaccinated individuals may not hold. Those feeling forced to vaccinate are less likely to do so in the future.

Session 10: Information | Room 2

Wednesday July 3, 2024 14:00 - 14:15

**Does ignorance love company? Social malleability of information avoidance and decision-making.**

Katharina Reher | Martin Götz | Filippo Toscano | Jörg Gross

University of Zurich | University of Zurich | University of Zurich | University of Zurich

Addressing challenges, like climate change, or migration crises, requires prosocial behavior. However, empirical research has shown that individuals voluntarily avoid information on the social impact of their decisions, when it serves their self-interest, thereby impeding prosocial behavior. Research has predominantly focused on the individual level, assuming personal preferences as the primary determinants of voluntary ignorance. Yet, research on cooperation has shown that prosociality can be highly sensitive to others' behaviors. In this pre-registered study, we investigated the group dynamics of information avoidance by testing the degree (1) to which individuals' decisions to remain ignorant are influenced by their peers' decisions, and (2) to which judgements about the appropriateness of voluntary ignorance are influenced by the prevalence of such choices.

In the baseline treatment, 60 individuals repeatedly decided whether to inform themselves about the possible negative consequences of their charitable donation decisions. Remaining ignorant can be attractive here, as it allows choosing the self-serving action without knowing the possibly negative social consequences. In two experimental treatments, participants were divided into groups of four. In the information treatment, 124 participants received information about others' decisions to avoid or disclose information after each round; in the choice treatment, 120 participants received feedback on others' choices for the selfish or (potentially more) prosocial option. Another 100 participants judged the social appropriateness of ignorance as third parties across different scenarios, varying how many people in a group remained ignorant.

Social influence had a significant effect on voluntary ignorance and normative judgement. First, group members' behavior in the previous rounds significantly affected their actions to remain ignorant and choose the selfish option. Second, voluntary ignorance was judged based on its frequency of occurrence. Interestingly, two clusters of people emerged in their normative judgements of ignorance: "Principled" raters judged the appropriateness of ignorance regardless of how widespread this behavior was in the group; "socially malleable" raters judged both, ignorance and disclosure of information, as more appropriate depending on how common it was. However, these clusters do not apply to following prosocial or self-serving actions; prosocial actions were predominantly judged as appropriate regardless of how common they were, whereas selfish acts were judged more leniently when they were more common.

The results highlight that voluntary ignorance and resulting selfishness are sensitive to social information. These findings underscore the importance of considering social influences and norms alongside individual preferences for further research on information avoidance of social decision consequences.

Session 10: Information | Room 2

Wednesday July 3, 2024 14:15 - 14:30

**Impact of Information Disclosure on Corruption and Cooperation: A Public Goods Game Approach**

Simone Righi | Francesca Pancotto | Francesca Federico

University of Modena and Reggio Emilia | University of Modena and Reggio Emilia | Arizona state university

Corruption significantly undermines public goods provision and societal welfare. This study explores how different levels of information disclosure about corrupt behavior influence corruption and cooperation levels in a Public Goods Game setting. The experiment extends the Bribe Game framework by Buffat & Senn (2019), introducing three different treatments which manipulate the information available to citizens to decide their behaviour and their behavioural options. The experimental design involves a baseline and three treatments with a between-subjects structure. In BG-FULL (reproducing Buffat & Senn 2019), all information about corruption is disclosed. BG-PINFO conceals to citizens the details about the corruption behaviour of other group members. BG-PENDO conceals the information but allows participants to buy access past corrupt activities. Finally, BG-PENDOEXIT further enables participants to opt out of the interaction after deciding on information purchase. The experiment was coded in Otree, and a total of 416 participants were recruited from the student population at the Ca'Foscari University of Venice. We run a total of 13 sessions, randomized and blinded treatment allocation, between October and November 2023. The sample size was determined – prior to the sessions - through power analysis, targeting meaningful detection of differences in corruption levels and public goods contributions across treatments. All details of the experiments were pre-registered in the OSF database.

The study hypothesizes that reducing the amount of information about corruption (i.e. comparing BG-FULL and BG-PINFO) or enabling endogenous information purchase (i.e. comparing BG-PENDO and BG-PENDOEXIT) reduces corruption and enhances contributions to the public good. The hypothesis posits that complete disclosure may misrepresent corrupt activities as rational behavior, fostering corruption through social learning. In contrast, limited or endogenously acquired information is expected to mitigate these effects by altering perceptions and control over corruption. Further the possibility of exiting from an undesirable situation is hypothesized to be conducive of more cooperation and less corruption, essentially by creating an outside option with respect to corruption. These hypotheses are tested both through correlational analysis and through regressions, including Treatment dummies, Pro-sociality preferences (measured through the Distribution Game (DG), gender, age, and educational background, perceived corruption in the general society and indicators of idiosyncratic social capital, which might influence how individuals behave in the game. Both tests return results that support the two posed hypotheses. The findings provide empirical insights into the role of information in corruption dynamics and offer guidance for policy formulations aimed at enhancing public goods provision and societal welfare.

Session 10: Information | Room 2

Wednesday July 3, 2024 14:30 - 14:45

**Giving (in) to help an identified other**

Linh Vu | Catherine Molho | Ivan Soraperra | Susann Fiedler | Shaul Shalvi

University of Amsterdam | Vrije Universiteit Amsterdam | Max Planck Institute for Human Development | Vienna University of Economics and Business | University of Amsterdam

Rationale: People give more to a person in need when this person's identity is known. Such altruistic behaviors may arise from a genuine concern for the person, leading people to give. Alternatively, altruistic behavior may also arise from one's attempt to keep a positive image and reduce the guilt of not giving, leading people to give in. In the current Registered Report, we tackle the question: Does facing an identified person lead people to give more or give in more?

Methods: In two preregistered studies (N = 3,671), participants made allocation decisions in transparent vs. ambiguous settings with a predetermined (versus undetermined; Study 1) or an identified (versus unidentified; Study 2) child in need as the recipient.

In the transparent setting, participants chose between two options: option A paid participants £0.6 and the child in need £0.1; option B paid both parties £0.5.

In the ambiguous setting, participants knew option A would always pay them more than option B, £0.6 instead of £0.5. They did not know, however, whether the child in need would receive £0.1 or £0.5 if they chose option A, as both possibilities were equally likely. Participants could resolve the ambiguity for free and learn the consequences for the child before making a choice or make a choice without the information.

Results: Consistent with our pre-registered hypothesis, results revealed that participants gave significantly less to undetermined/unidentified children in an ambiguous, compared with a transparent setting (H2). However, in contrast to our predictions, predetermined/identified children did not receive more than undetermined/unidentified children in transparent settings in which they know how their choice impacts the children (H1). Accordingly, the predicted interaction between identification and ambiguity was not significant (H3). Participants' likelihood of resolving the ambiguity was independent of the child's identification (H4). However, exploratory analyses revealed that participants who willingly resolve the ambiguity surrounding the impact of their choice gave more compared to those who were given transparent information by default.

Conclusion: Overall, the results suggest that some people give in when making their donation decisions. However, the tendency to give in, to preserve a positive image, is independent of whether the recipient is identified or not. Our findings suggest that to increase donations, designing a transparent environment, in which potential donors receive explicit information about how their actions impact others, would be valuable for charity organizations.

Session 10: Information | Room 2

Wednesday July 3, 2024 14:45 - 15:00

**Facilitating Cooperation by Manipulating Attention**

Claire Lugrin | Arkady Konovalov | Christian C. Ruff

University of Zurich | University of Birmingham | University of Zurich

Rationale: Cooperation is essential for human societies, but not all people cooperate to the same degree. While these individual differences are usually explained by motives such as other-regarding or risk preferences, at least some differences might also relate to how people attend to the choice-relevant information. Here we study with eye-tracking and computational modelling how cooperation is linked to attentional mechanisms.

Methods: 84 subjects played 192 trials of a one-shot prisoner's dilemma (PD) game with an anonymous real opponent while their eye movements were recorded. On each trial, both players chose to either cooperate or defect. While the players' dominant strategy is to always defect, mutual cooperation leads to a higher total payoff than mutual defection. By systematically varying the four payoff values, we created 96 unique PD games, presented in a standard matrix form and randomly varied the position of the options (cooperate or defect) for both the subject (columns) and the opponent (rows) on each trial. Subjects received no feedback except for one randomly selected game paid out at the end of the session. We fit a computational utility model to subjects' choices and used mixed-effects logistic regressions and machine learning models to estimate the interacting effects of game payoffs, information position, and gaze on the subjects' decisions.

Results: As expected, cooperative behavior was influenced by the payoff values ($p < 0.013$) and was linked to other-regarding preferences ($R2 = 0.74$, $p < 0.001$). However, cooperation was also linked to the relative attention directed to different payoffs: Classifiers trained on gaze sequences were able to accurately predict out-of-sample cooperation (predicted vs measured cooperation rate $R2 = 0.8$, $p < 0.001$), showing the importance of information sampling mechanisms. Manipulations of the payoff locations significantly affected cooperation rates (row orders $p < 0.001$, interactions of rows and columns orders $p = 0.045$), but not estimated other-regarding or risk preferences ($p > 0.5$). Importantly, the location manipulations altered subjects' gaze behavior and affected the first payoff subjects attended to, exogenously driving their attention to certain gaze sequences favoring higher cooperation rates.

Conclusions. Our results suggest that cooperation does not only depend on payoffs and preferences, but also on attentional mechanisms that can be manipulated exogenously. This has implications for our understanding of individual differences in cooperative behavior and suggests attentional interventions that could enhance cooperation.

Session 10: Information | Room 2

Wednesday July 3, 2024 15:00 - 15:15

**Comprehension in Economic Games**

Lina Koppel | David Andersson | Magnus Johannesson | Eirik Strømland | Gustav Tinghög

Linköping University | Linköping University | Stockholm School of Economics | Western Norway University of Applied Sciences | Linköping University

Most disciplines rely on economic games to measure prosocial behavior in controlled experimental settings. However, if participants fail to understand a game's payoff structure, it is difficult to infer their underlying motives from their behavior. For example, a selfish individual confused about the game's incentives may be erroneously classified as having social preferences. Here, we investigate participants' comprehension of the payoff structure of five standard economic games commonly used to study social preferences: the Dictator Game, Ultimatum Game, Trust Game, Public Goods Game, and Prisoner's Dilemma.

Participants were recruited from two online platforms, CloudResearch (using the MTurk toolkit; n = 540) and Prolific (n = 528), as well as a student subject pool in a laboratory setting (n = 500). Participants first played each of the five games (in random order), and then completed a comprehension check for each of the games. Methods and data analysis plans were preregistered on OSF.

In the online samples, the game with the highest level of misunderstanding was the Trust Game (with 70% of participants answering incorrectly on one or more comprehension questions), followed by the Public Goods Game and Prisoner's Dilemma (each at 52%), the Ultimatum Game (27%), and the Dictator Game (24%). In the lab sample, the game with the highest level of misunderstanding was the Public Goods Game (53%), followed by the Trust Game (47%), Prisoner's Dilemma (30%), Ultimatum Game (25%), and Dictator Game (22%). Preregistered logistic regressions indicated that: (1) misunderstanding was greater on Prolific than CloudResearch in the Dictator Game, Ultimatum Game, and Public Goods Game (ps < .005; differences ranging from 14 to 26 percentage points); (2) misunderstanding was greater online than in the lab in the Trust Game (by 20–23 percentage points, p < .005) and Prisoner's Dilemma (by 13–17 percentage points; p < .005); (3) incentivizing the comprehension check had no statistically significant effect on misunderstanding in any of the games or samples (all ps > .05); and (4) higher numeracy predicted lower misunderstanding in all games and samples (ps < .005). Additionally, misunderstanding predicted increased prosocial behavior in several of the games, although the effect and strength of evidence varied across models.

Taken together, our findings suggest that misunderstanding may be an important factor in explaining prosocial behavior, and that reliance on standard one-shot games may lead researchers to overestimate the importance of social preferences.

Session 10: Information | Room 2

Wednesday July 3, 2024 15:15 - 15:30

**Negotiators have the wrong model of their counterparts: What really motivates negotiators' behavior?**

Shira Garber-Lachish | Simone Moran | Boaz Keysar | Yoella Bereby-Meyer

Ben-Gurion University of the Negev | Ben-Gurion University of the Negev | University of Chicago | Ben-Gurion University of the Negev

Negotiators face two dilemmas: Whether to be honest and whether to believe their counterparts. We suggest that negotiators apply different motivation models for the two. When deciding whether, to be honest, negotiators are motivated not only by self-gain but also by moral-image concerns and anticipated guilt. When contemplating counterparts' honesty, they assume that others are primarily motivated by self-gain, and therefore underestimate their counterparts' honesty. In five studies (Total N = 1,717), participants were assigned to either a Behavior condition, playing the role of sellers of a faulty device, and incentivized for the deal they closed, or to an Expectation condition, incentivized to accurately predict sellers' disclosure of the malfunction.

Study 1: Behavior participants chose between a truthful (conveying malfunction) or deceptive (concealing malfunction) opening message. Expectation participants estimated which message the seller would choose. The rate of sellers choosing the truthful message was significantly higher than estimated.

Study 2 (https://aspredicted.org/4C2_F4J): We measured whether or not sellers mentioned the malfunction during a free negotiation chat. We assigned expectation participants to the role of buyers and varied the expectation framing by asking them to estimate if sellers would mention the malfunction (be honest) or not (be deceptive).

Although buyers were somewhat more trusting when predicting honesty versus deception, they significantly underestimated sellers' actual honesty under both framings.

Study 3: We manipulated the extent to which the negotiation evoked moral concerns by varying whether sellers were prompted with a direct question about the malfunction. Answering direct questions dishonestly involves lies of commission, which increase moral-image threat and anticipated guilt. Consistent with our theorizing that negotiators' honest behavior (but not expectations) is sensitive to moral concerns, the actual proportion of honest seller responses was higher in the direct versus general question condition, yet the proportion of buyers predicting honest seller responses didn't vary by question type.

Study 4 (https://aspredicted.org/LX1_NMH) provided additional support for our theory by finding that the proportion of buyers expecting honest sellers significantly increased when prompting buyers to consider sellers' moral concerns before (versus after) estimating sellers' honesty. Notably, however, in both cases, the expected honesty remained lower than the actual honesty.

Our findings enhance the understanding of distrust in negotiation. We show that people tend to underestimate other negotiators' honesty due to not realizing that even in negotiations, people

have moral concerns. Practically, findings shed light on ways to enhance negotiation trust, potentially impacting consequent negotiation processes and agreements.

Session 11: Inequality & Fairness | Room 1

Thursday July 4, 2024 09:00 - 09:15

**The rich, the poor and strength of Inequality: evidence from a meta-dataset of public good games experiments.**

Rémi Suchon | Vincent Théroude

ANTHROPO LAB – ETHICS EA 7446, Université Catholique de Lille, | Université de Lorraine

Rational: Cooperation is at the root of a great many economic activities. At the level of a country, social capital, the ease with which strangers cooperate, has a sizable economic payoff (Knack and Keefer, 1997). At the micro level, cooperation is often necessary to achieve economic efficiency. For instance, community members must cooperate to avoid the exhaustion of common resources. Given the well-documented recent increase in inequalities across the world (e.g., Piketty 2014), a pressing question is whether cooperation can be sustained between agents with unequal resources. The effect of inequality on cooperation is theoretically uncertain, empirically hard to identify with observational data, and has important policy implications.

Method: To contribute to this question, we identified all the published economics experiments in which heterogeneous endowments are introduced into a linear public good game. We collected the original dataset from most of these studies to build a meta-dataset, which includes 24 papers, more than 100 treatments, and 60000 observations at the individual level. We coded a wealth of relevant variables at the paper, treatment and group level. Notably, for each observation, we computed the Gini index of endowments. The Gini index is a well-known measure of inequality. It is continuous, allowing us to measure not only the effect of the presence of inequality, but also of the strength of inequality.

Results and conclusion: We run two complementary analyses. First, we investigate the effect of inequality on contributions at the group level, i.e., we relate our measure of inequality to the share of the sum of endowments contributed within a group. We document a significant negative effect of the strength of inequality on cooperation that cannot be fully due to publication bias. Second, we identify that the negative effect exists at the intensive margin: it is not only a matter of the presence of inequality but also of its strength.

Second, we investigate individual contribution decisions. We identify a contribution gap between the rich (those whose endowments are higher than the median in their group) and the poor: the rich contribute a smaller share of their endowments than the poor, or than the members of equal groups. We also find that this gap increases as the strength of inequality (measured by the Gini coefficient of endowment) increases.

Therefore, we conclude that the negative marginal effect of inequality on cooperation is driven mainly by the “rich” who reduce their participation in the public good as inequality increases.

Session 11: Inequality & Fairness | Room 1

Thursday July 4, 2024 09:15 - 09:30

**Moving up? The effect of economic mobility on giving behavior**

Leticia Micheli

Leiden University

Perceptions of economic mobility, that is, perceptions of the possibility of changing socioeconomic classes in society, have been shown to influence a range of individual attitudes and behaviors, such as support for redistribution, gambling, status-seeking and materialism. However, it is less clear how perceptions of economic mobility may influence interpersonal behavior. Recent studies have demonstrated a positive effect of perceptions of upward economic mobility on self-reported prosocial behavior and intentions to donate money or time. Yet, these studies did not consider that, in unequal societies, prosocial behavior might vary according to the recipient of the generous action.

This study investigates whether economic mobility influences giving behavior when the recipient is someone in ranks below or above one's own. Importantly, giving behavior is incentivized.

The sample included 460 participants, predominantly from Global South countries (recruited via Besample). Participants were randomly assigned to either a low or high rank on a 5-rank ladder, resembling socioeconomic classes. To manipulate economic mobility, participants played a game in which they estimated the number of dots in a picture. Participants assigned to the high mobility condition earned 3 points per correct estimation, while those in the low mobility condition earned 1 point per correct estimation, making it considerably harder for them to accumulate points and move up the ranks. Participants then played 5 one-shot decisions in the dictator game, each paired with individuals in one of the different ranks of the ladder.

Although both the manipulation of rank and economic mobility were successful, we found no effects of economic mobility or ranks on overall giving behavior. We also found no effects of mobility on giving behavior when the recipient belonged to ranks below or above the participants' own ranks. These results contradict previous findings showing a positive effect of subjective perceptions of mobility on self-reported prosocial behavior. Our results could indicate that, when using incentivized measures, economic mobility does not influence giving behavior. Alternatively, our results could be particular to a Global South sample, which was recently found to be less meritocratic than Global North countries. A replication of the present study in Global North countries is underway and results will be discussed during the conference.

Session 11: Inequality & Fairness | Room 1

Thursday July 4, 2024 09:30 - 09:45

**The Impact of Economic Inequality on Charitable Behavior**

Jing Lin | Yu Kou

Beijing Normal University | Beijing Normal University

Previous research suggests that economic inequality has caused a wide range of negative societal impacts. However, little is known about how economic inequality influences charitable behaviour as a socioecological environment determinant. Study 1 utilized the World Giving Index report to examine the relationship between objective economic inequality (i.e., Gini coefficient) and the proportion of charitable donors across 165 countries/regions. In Study 2, we used extensive data from five waves of the China Family Panel Studies (N = 66,759) to explore the negative association between objective economic inequality and monetary donations within China. This study extends the investigation to a national context, offering insights into how economic disparities at a domestic level impact charitable giving. Study 3 employed a computational model to observe the interactive differences in donation behavior among agents under varying levels of inequality. By analyzing a broad range of interaction samples, this study investigates the dynamic interaction between objective economic inequality and donation behaviors. Studies 4a and 4b focus on the micro-level individual perceptions and behaviors. Study 4a, through an online survey (N = 243), examined the correlation between perceived economic inequality and charitable donations and volunteer services, testing the mediating role of status anxiety. Study 4b manipulated the level of perceived economic inequality through recollection tasks in a laboratory setting, followed by measuring status anxiety and responses to online charitable donation requests. This approach not only explores the causal relationship but also reaffirms the mediating role of status anxiety. This research, against the backdrop of widening wealth gaps, reveals how a macro-socioecological factor like economic inequality impacts individual behavior from both macro-societal and micro-individual perspectives. It enriches the understanding of the mechanisms through which economic inequality influences behavior. Given the critical role of donations in mitigating wealth disparities and maintaining social stability, our findings also provide scientific recommendations for social governance and fostering positive social mindsets in societies with significant wealth gaps.

Session 11: Inequality & Fairness | Room 1

Thursday July 4, 2024 09:45 - 10:00

**Pro-sociality of the financially affluent and deprived**

Leon P Hilbert | Marret K Noordewier | Wilco W Van Dijk | Angelo Romano

University of Amsterdam | Leiden University | Leiden University | Leiden University

There is still no consensus in the current literature whether someone's own financial situation (e.g., being poor vs rich) affects their prosociality. In addition, it is unclear whether and how this prosociality might depend on the recipients financial situation, and how people might decide to signal or hide their own neediness or privilege.

To shed further light on these open questions, we conducted an online experiment where we manipulated participants available resources and measured their prosociality. Participants first completed an incentivized task where they had to manage the finances of a household. Over multiple rounds, participants earned income by doing a task and paid regular expenses. Between conditions, we varied the height of the expenses, such that some participants accumulated debts while others accumulated savings. Participants then were asked to indicate their strategy for an incentivized one-shot dictator game. As senders, they were asked how much of an additional endowment they would keep (which would be added to their own balance) or send to another participant. Here, participants made separate sending decisions for recipients with savings, debts, and unknown financial situation. In addition, participants indicated as recipients whether they would signal their own financial situation (i.e., debts or savings) to the sender, separately for each sender types. After completion of data collection, participants were randomly matched in sender and recipient pairs and the respective strategy was executed for their incentivized payment.

Results showed that, across recipient types, participants own financial situation did not predict their prosociality. At the same time, senders with debts were less prosocial towards recipients with savings, but not towards recipients with debts. Moreover, regardless of their own financial situation, senders were more pro-social towards recipients with debts than recipients with unknown financial situation and recipients with savings. The latter two recipient types also differed in received allocations, such that senders were less prosocial towards recipients with savings than towards recipients with unknown financial situation. Last, signaling decisions showed that recipients with saving correctly anticipated this pattern. They strategically hid their own wealth particularly from senders with debts, thereby increasing their own payoff.

These findings were largely in line with predictions based on self-interest or inequality aversion, and less so with predictions based on either in-group preferences or scarcity theory.

**People prefer to address inequalities by reducing disadvantage over advantage**

David Munguia Gomez | Daniela Goya-Tocchetto | Megan Burns

Yale School of Management | University at Buffalo | Yale School of Management

In Study 1 (Prolific, N=129), participants indicated their support for solutions that addresses disadvantages or advantages relating to various inequalities. To avoid cherry-picking solutions that would bias the test of our hypothesis, we relied entirely on ChatGPT-4 to generate the stimuli. ChatGPT created a list of 10 inequalities, each with two solutions—one for an advantaging mechanism and one for a disadvantaging mechanism, making 20 solutions. For each inequality, participants were randomly assigned to rate either the advantage or disadvantage solution for each inequality. Overall, participants were more supportive of a solution when it addressed a disadvantaging mechanism than an advantaging one ($b = 0.19$, $p = .066$; regression predicting support based on the type of solution [disadvantage, advantage] and inequality seen, clustering errors at the participant level). Moreover, for seven out of the ten inequalities, participants indicated greater support for the solution focused on disadvantage than advantage.

In Study 2 (Prolific, N=593), participants evaluated a proposal to create more equality in a college context, which would lead to admitting 5 lower-income applicants and rejecting 5-higher-income applicants. We framed the proposal as either reducing a disadvantage ("lowering the admissions bar for lower-income applicants"), reducing an advantage ("raising the admission bar for higher-income applicants"), or simply as leading to the outcome. To ensure that participants noticed the framing and identical outcome across the framings, we presented them with a short video of each proposal. Participants were more supportive of the proposal when it was framed as reducing a disadvantage than an advantage, both between-subjects ($t[416] = 3.10$, $p = .002$) and within-subjects ($b = -0.42$, $p < .001$, regression predicting support based on framing [disadvantage, advantage] and order, clustering errors at the participant level). Support for the neutral, outcome-only framing was in between the other two conditions and not significantly different from either ($p = .197, .101$).

Our findings suggest that people are more inclined to support policies aimed at reducing disadvantages rather than diminishing advantages, even when the end results are identical. Efforts to reduce inequality may be held back by a preference for actions that level the playing field for the underprivileged over those that limit the privileged.

**Affirmative Action: Within-group Inequality in Competitive Environments**

Ankush Asri | Urs Fischbacher | Jan Hausfeld | Yvette Lambi

Radboud University | University of Konstanz | University of Amsterdam | University of Amsterdam

Inequalities are perpetuated through mechanisms that disadvantage certain groups and advantage others. While addressing inequalities can be achieved both by reducing disadvantages and advantages, across two experiments we find that people are more willing to do so by targeting disadvantage.

Inequality in wealth, income, and opportunities based on natural identities such as race and gender are widespread. One policy tool to ”level the playing field” is Affirmative Action (AA) policies. Previous research has shown that AA policies effectively increase disadvantaged groups’ willingness to compete. However, most studies focus on resource heterogeneity between groups, even though significant differences exist within groups. This study investigates the effect of the group- and income-based AA policies on willingness to compete, fairness perceptions about the AA policies, and intergroup relations with both within- and between-group inequality. We conducted an online experiment where participants were randomly assigned to a rich and a poor group. We run two treatments, with within-group homogeneity and within-group heterogeneity in endowments. We focus on three institutional setups for selecting tournament winners: no-AA, group-based-AA (at least one poor group winner irrespective of own endowment), and income-based-AA (at least one poor winner regardless of the group).

In line with previous literature, we find that the willingness to compete increases for beneficiaries and does not decrease for non-beneficiaries with within-group homogeneity. However, within-group heterogeneity discourages non-beneficiaries, leading to no change in overall investments, which is missing from the previous literature. We further find that the out-group is punished more, and low-endowment participants receive more bonus points, irrespective of the policy. We also find that no-AA is the fairest for everyone. However, the choice as a social planner leans towards “own benefit” for both beneficiaries and non-beneficiaries. Apart from contributing to the academic literature addressing an essential question on inequality within groups, we provide evidence on why AA institutions should be chosen more carefully, considering the pre-existing inequalities within groups and not just between groups based on natural identities.

Session 12: Reputation & Communication | Room 2

Thursday July 4, 2024 09:00 - 09:15

**The Costs and Benefits of Gossip**

Terence Daniel Dores Cruz | Kim Peters | Romy van der Lee | Bianca Beersma

University of Amsterdam | University of Exeter | Vrije Universiteit Amsterdam | Vrije Universiteit Amsterdam

Study Rationale: Gossip is integral to cooperation because it is the building block of reputation-systems that entail reputational costs and benefits for prosocial and antisocial behavior. Despite this, gossip is commonly perceived negatively. This presents a puzzle: Gossip is needed for cooperation but could lead to costs for gossipers that motivate refraining from gossip. Moreover, for gossip to support reputation-systems, it is essential that true/prosocial gossip is associated with benefits while gossip that harms reputation-systems should be associated with costs. Shedding light on gossip's consequences builds our understanding of how conditions for reputation-based cooperation are met.

Methods: Across 3 studies, potential gossipers (trustors) interact in a trust game with gossip targets (trustees). Potential gossipers subsequently can send gossip to a receiver that will interact with the target next. Receivers then also reward or punish potential gossipers in a dictator game with giving and taking. In two studies pre-registered studies, we first tested how receivers respond to true or false and positive or negative gossip as well as no (gossip) communication across 6 (Study 1: n=201; nobs=1206) and 8 rounds (Study 2: n = 200; nobs = 1600). We are running a pre-registered interactive laboratory study to focus on sending gossip (Study 3; planned n=300; nobs=7200; 12 rounds including the dictator game, 12 rounds without).

Results: Studies 1 and 2 showed that when senders shared true gossip that contributed to the reputation-system, this was associated with receivers providing more benefits, or less costs, to senders as compared to when senders did not contribute to the reputation-system by sharing false gossip, or not communicating gossip. Moreover, within true gossip, positive gossip was associated with more benefits than negative gossip. From the receiver's perspective, this indicates that there are fewer costs to gossiping than refraining from gossip. Yet, this provides no insight into how potential gossipers perceive the costs of gossip or how this informs their behavior, which will come from Study 3.

Conclusion: The first steps towards understanding the costs and benefits of gossip show that receivers' behavior follows predictions based on gossip being important for reputation-based cooperation. Harming the reputation-system by not contributing gossip or spreading false information is punished while accurate gossip contributions are rewarded – especially when positive. With our test of gossiper behavior, we will provide further insights into whether these social costs of gossip could hinder or promote gossip and build our understanding of how reputation-based cooperation functions.

Session 12: Reputation & Communication | Room 2

Thursday July 4, 2024 09:15 - 09:30

**Partner Perceptions During Brief Online Interactions Shape Partner Selection and Cooperation**

Tiffany Matej Hrkalovic | Bernd Dudzik | Hayley Hung | Daniel Balliet

Free University Amsterdam & Delft University of Technology | Delft University of Technology | Delft University of Technology | Free University Amsterdam

Today's society is characterized with a certain level of interdependence, where humans require assistance from other individuals on a daily basis. How people select partners in these situations can affect their individual and mutual outcomes in the near or distant future. Despite the theoretical and experimental propositions of the relevance of partner selection, there is still a gap in the literature trying to understand the complex relationship between nonverbal behavioral cues, person perceptions, partner selection and cooperation in semi-naturalistic settings.

Thus, the aim of this research was to investigate: (a) the role of situational affordances and person perceptions in informing partner selection for a cooperative task, (b) whether people are accurate in their perceptions of others' (task-relevant) characteristics, (c) the role of person perceptions and partner selection in predicting cooperative behavior, and finally, (d) whether we could use non-verbal behavioral cues to predict the perceptions of a potential partner's characteristics during a brief online interaction.

Here, individuals participated in an interactive study with the goal of selecting a partner for an interdependent task that either afforded for the expression of warmth (Joint Trust Task) or competence (Joint Competence Task). Participants had a 3-minute (online) conversation with up to five individuals, reported their evaluations, selected partners for the specific task and engaged in the specific task with each participant, respectfully.

Using machine learning, an exploratory analysis found that facial expressiveness and acoustic cues captured during conversations were predictive of warmth perceptions, but not competence. While competence was not predictive by nonverbal behavioral cues, both warmth and competence perceptions were predictive of whom participants selected for partners. Each trait was more predictive of partner selection within the task that afforded for that trait. Lastly, partner selection had a role in informing participant's cooperative behavior, where participants were more cooperative towards selected, compared to unselected partners. However, we found that perceptions formed during interactions were not predictive of the chosen partners' actual cooperative behavior, raising doubts about the accuracy of person perceptions informing partner selection. All analyses were preregistered using OSF, except the machine learning experiments, which was due to the novelty of the RQ.

These findings raise interesting questions about the effectiveness of people's abilities in partner selection and the contextual variables that need to be considered for its understanding. Furthermore, study's insights offer valuable implications for improving decision-support systems aimed at aiding individuals in making more informed choices when selecting future collaborators.

Session 12: Reputation & Communication | Room 2

Thursday July 4, 2024 09:30 - 09:45

**Words are not Wind - How Joint Commitment and Reputation Solve Social Dilemma**

Marcus Krellner | The Anh Han

University of St. Andrews, UK | Teesside University, UK

Rational: The concept of joint commitment plays an essential role in shaping our social world and separating us from other primates. The essence of 'joint' lies in the mutual agreement that both parties refrain from making promises unless the other party does the same. When we need to coordinate for the best mutual outcome, any commitment is beneficial. However, when we are tempted to free-ride (i.e. in social dilemmas), commitment serves no obvious purpose. We show that a reputation system, which judges action in social dilemmas only after joint commitment, can prevent free-riding.

Keeping commitments builds trust. We can selectively enter joint commitments with trustworthy individuals to ensure their cooperation (since they will now be judged). We simply do not commit to cooperate with those we do not trust, and hence can freely defect without losing the trust of others. This principle might be the reason for pointedly public joint commitments, such as marriage. It is especially relevant to our evolutionary past, in which no mechanisms existed to enforce commitments reliably and impartially (e.g. via a powerful and accountable government).

Methods: Using evolutionary game theory, we study a model akin to preceding studies on indirect reciprocity, in which players can cooperate conditionally on the reputation of the players they meet. Building upon the latest advancements in this field, we consider that every player has their own private opinions about every other player. We analytically predict average reputation values, using them to calculate payoffs. These payoffs serve as the basis for simulating evolutionary dynamics through Monte Carlo simulations.

Results: Although uncooperative strategies sometimes invade, the predominant trend reveals that players who commit exclusively when trust is established and cooperate solely within the context of joint commitments emerge as the most prevalent. Following this, there exists a secondary group characterized by a mix of somewhat more trusting committers and unconditionally cooperative players.

Conclusions: These findings are notable, because much research from anthropology, philosophy, and psychology made the assumption that past collaborations were mutually beneficial and had little possibilities to free-ride. Our approach, grounded in evolutionary game theory, demonstrates that such assumptions are not obligatory, because free-riding could have been dealt with joint commitments and reputation. This reevaluation prompts a more nuanced understanding of the dynamics underlying cooperative efforts in human history.

Session 12: Reputation & Communication | Room 2

Thursday July 4, 2024 09:45 - 10:00

**Cooperation in the 'Helping Game': Good Standing or Image Scoring?**

Andrea Marietta Leina

University of East Anglia

In a lab experiment, we assess the effectiveness of reputation-based mechanisms in the decision to help strangers within the 'helping game,' a large group setting where cooperation is socially desirable. We compare 'good standing' (GS), a binary score with a recursive feature, and 'image scoring' (IS), a numerical score considering past actions. Two primary research questions guide my investigation: (i) In a homogeneous cost game, does GS better support indirect reciprocity than the IS mechanism? (ii) How do these mechanisms induce reciprocity in a heterogeneous cost game with two costs? We theoretically argue that the 'good standing' mechanism creates an incentive to discriminate between subjects who do not help someone that did not help in the past (‘justified punishers’) and subjects who did not help without a justification (‘unjustified non-helpers’). Analyzing data from our lab experiment with 144 participants, we find that GS effectively discriminates between subjects, supporting reciprocal helping, particularly in the heterogeneous cost game. However, IS leads to higher cooperation rates, regardless of the subject's reputation in both game scenarios.

Session 12: Reputation & Communication | Room 2

Thursday July 4, 2024 10:00 - 10:15

**The dynamics of universal cooperation with reputations**

Jacobus Martin Smit | Fernando P. Santos

University of Amsterdam | University of Amsterdam

In many social dilemmas, cooperation is not a rational strategy from an individual perspective despite being socially optimal. Indirect reciprocity (IR) has been proposed as a key mechanism of explaining the origin of human cooperation. Under IR, actions taken by interacting individuals are judged by third-party observers, considering the context in which actions were taken. Previous works have derived which reputation update rules (i.e. social norms) are more able to sustain long-term cooperation. A relatively unexplored line of research involves examining the interplay between reputation-based cooperation (under IR) and in/out-group conditioned cooperation. Empirically, it has been shown that reputations can overcome biases rooted in group-identity, such that universal cooperation can be sustained with IR [1]. Theoretically, however, the formal conditions for cooperation to be evolutionarily stable in group-structured populations remain unknown in scenarios where information about both reputations and group-identity are available.

In this presentation, we will describe some recent theoretical work exploring the stability of fair (i.e., reputation-based) and unfair (i.e., group-based) cooperation in group-structured populations. By advancing models that explicitly consider individuals belonging to different groups, we can also explicitly consider social norms that judge actions based on whether an interaction occurred between members of the same group or a different one, constituting an in-group or out-group interaction respectively. We study how unfairness may arise in populations where groups are unequally sized, even if the agents themselves are identical. We will present models based on social learning, which can be analysed with tools from evolutionary game theory, and models based on trial-and-error individual learning, studied through reinforcement learning methods [2].

We observe that a defecting majority leads a minority to defect, but not the inverse. Moreover, controlling the norms to judge in- and out-group interactions can steer a system towards either fair or unfair cooperation. Beyond equilibria analysis, the need to judiciously set norms to reach fair cooperation is clearer with independent RL agents, where convergence to fair cooperation occurs with a narrower set of norms. Our results highlight that, in heterogeneous populations with reputations, carefully defining interaction norms is fundamental to tackle both cooperation and fairness dilemmas.

[1] Romano, A., Balliet, D., & Wu, J. (2017). Unbounded indirect reciprocity: Is reputation-based cooperation bounded by group membership?. Journal of Experimental Social Psychology, 71, 59-67.

[2] Smit, J., & Santos, F. P. (2023). Learning Fair Cooperation in Systems of Indirect Reciprocity. In Adaptive Learning Agents Workshop (ALA @ AAMAS 2023)

Session 12: Reputation & Communication | Room 2

Thursday July 4, 2024 10:15 - 10:30

**Conformity versus credibility: A coupled rumor-belief model**

Wei Zhang

ETH Zürich

The opinion individuals hold about circulating information moderates their spreading behavior, as they may choose to spread rumors they believe in and remain silent about those they do not. Meanwhile, these opinions may be influenced by peers' opinions and observed behaviors.

In this work, we present coupled dynamics that integrate opinion formation and information spreading within two-layer multiplex social networks. In an influence layer, opinions are shaped by information credibility and social influence, the degree of which is determined by individual tendencies to conform. In a communication layer, information spreads conditioned on whether individuals believe in it.

Our game-theoretic model posits that individuals form their opinions based on observed spreading (or sharing) from others. a social tendency to conform with their peers' opinions, and an additional factor: information credibility.

For example, when a piece of information or an opinion is more credible than another, individuals with an appreciation for the more credible option are less likely to conform to their peers' less credible opinions. Such a situation is rarely addressed in existing theoretical frameworks.

Using analytical and numerical tools, we demonstrate consistent behavior of the model by identifying the criteria for an evolutionary advantage in opinion formation and a critical spreading rate for rumor outbreaks. We then proceeded to examine two scenarios distinguished by considering credibility as either an intrinsic or a social property of information. Our results show that, for intrinsic credibility, higher degrees of conformity promote the spread of incredible information and inhibit the diffusion of credible information; for social credibility, higher degrees of conformity have an amplifying effect on the evolution of belief and rumor spreading. Additionally, the former scenario exhibits rich behaviors in the interplay between opinion evolution and information spreading, including paradigm shift, passive support, and fringe belief. We thus contribute to the understanding of the dynamics of opinion formation and information propagation in social groups. Empirical application may inform efforts to assess and combat the spread of misinformation.

The work is published in Chaos, Solitons & Fractals, 2023. https://doi.org/10.1016/j.chaos.2023.114172

Session 13: Ingroup-Outgroup | Room 1

Thursday July 4, 2024 11:00 - 11:15

**The role of strategic uncertainty in collective risk social dilemmas with donors**

Natalie Struwe | Ivo Steimanis | Esther Blanco | Julian Benda

University of Innsbruck | Philipps University Marburg | University of Innsbruck | University of Innsbruck

This study addresses the role of strategic uncertainty in explaining previous pessimistic results for collective risk social dilemmas. Experimental evidence in Gross & De Dreu (2019, SciAdv) and Gross & Böhm (2020, PNAS) suggests that when groups face a common threat, individuals resort to private protection over collective avoidance of the threat, if given the opportunity. We show that a large proportion of this overwhelming self-reliance can be attributed to strategic uncertainty about others' behavior rather than selfishness.

We expand the collective risk social dilemma to incorporate a subgroup ("outsiders") who are affected by the collective risk but do not have the capacity to protect against it. A second subgroup ("insiders") can invest into a collective solution (protecting all group members through a threshold public good) or a less efficient but safe individual solution (protecting only themselves), as introduced in Gross & Dreu (2019). Our novel design relates to field contexts where parts of the population have different capacities to protect against collective damages associated with, for example, losses of ecosystems and occurrence of natural disasters.

In a pre-registered laboratory experiment (https://aspredicted.org/SXY_YVF) with 400 participants (sample size determined through a power analysis), we consider a repeated decision-environment with four treatment conditions, systematically varying strategic uncertainty: i) a baseline scenario with passive outsiders, ii) proportional arrangements where outsiders can send donations to compensate insiders for collective solution efforts, thereby changing the Nash Equilibrium, iii) pledges allowing for non-binding commitments among both subgroups, and iv) reducing the problem to a single insider and single outsider. We hypothesize that all treatments with transfers will increase the likelihood of avoiding collective damages compared to the baseline scenario, with the single insider-outsider treatment resulting in the highest efficiency to avoid collective damages.

Our results are based on multilevel regressions for difference-in-differences estimates (accounting for pre-treatment group-specific variation). We find that all three treatments significantly improve the proportion of collective solutions, with more protected outsiders and less resources lost. Further, removing strategic uncertainty with a single insider and outsider entails a higher likelihood of avoiding collective damages than a group of insiders and passive outsiders. Most importantly, our findings suggest that compared to the setting with passive outsiders, institutions involving proportional transfers to the group of public good providers can increase the avoidance of collective damages to the same extent as removing strategic uncertainty with a single public good provider and a single outsider.

**Not all groups are created equal: Unequal abilities to cooperate within groups increases inequality and decreases group-transcending cooperation**

Filippo Toscano | Katharina Reher | Martin Götz | Jörg Gross

University of Zurich | University of Zurich | University of Zurich | University of Zurich

Cooperation can manifest at various levels of social organization—between individuals, groups, or larger collectives. At the group level, cooperation enables individuals to generate shared benefits that every member can enjoy. Importantly, groups can differ in their effectiveness to cooperate. Some groups may have better means of creating shared benefits for themselves than others. Such between-group inequality can arise, for example, between countries with different socio-economic development, creating a misalignment in the incentives to cooperate within vs. across group boundaries. Here, we empirically investigate the repercussions of this misalignment regarding cooperation, inequality, and wealth creation.

Using a nested social dilemma, we divided 378 participants into two groups of three within a larger collective (n = 63). Over 20 rounds, participants could allocate resources in self-serving ways (keeping resources), for the benefit of their group only ('group cooperation'), or for the benefit of everyone, irrespective of group membership ('universal cooperation'). In two control treatments, we manipulated the attractiveness of group cooperation by offering either a high or a low return for group cooperation compared with universal cooperation.

As expected, when group cooperation yielded a low return, the groups exhibited greater cooperation across boundaries. Conversely, when group cooperation provided a higher return, more group-exclusive, cooperation emerged. To introduce misalignment, in our experimental treatment we assigned one group a high ('high club good') and the other a low ('low club good') return for group cooperation. Compared with the control treatments, groups with a high club good displayed more universal cooperation, showing concern for the other group; in contrast, groups with a low club good cooperated less universally and free-rode more, potentially anticipating that universal cooperation would not be reciprocated accordingly by the other group. In fact, groups with a low club good still cooperated more universally than groups with a high club good. As a result, earnings inequality between groups increased, and those with a high club good benefited from the introduced between-group inequality. On the other hand, those with a low club good earned less than their counterparts in the control treatment.

In conclusion, disparities in the attractiveness of group cooperation appear to be noticed and acted upon, but not strongly enough to lead to equality through (reciprocal) universal cooperation. On the contrary, this disparity intensified the social dilemma and exacerbated wealth inequality between the groups, highlighting important novel challenges to intergroup cooperation.

Session 13: Ingroup-Outgroup | Room 1

Thursday July 4, 2024 11:30 - 11:45

**Groupiness over time: A longitudinal lab-in-the-field experiment**

Hannes Rusch

Max-Planck-Institute for the Study of Crime, Security and Law

Study rationale: Opportunities in life de facto hinge on group memberships that de jure should not matter for individual outcomes. But why do racism and other forms of discrimination endure in our societies? Kranton et al. 2020 (PNAS 117/35) recently suggested that people may differ systematically in their responses to markers of group membership: highly 'groupy' people readily discriminate based on any sort of group marker, while 'non-groupy' people tend to disregard group memberships of others. Here, we ask: if Kranton et al.'s findings replicate, (i) is 'groupiness' better conceived of as a trait or as a state, and (ii) can we identify possible correlates of 'groupiness' at situational- or individual-levels?

Methods: In a preregistered, longitudinal lab-in-the-field experiment (N=141), we assessed individuals' 'groupiness' using the same set of incentivized economic games like Kranton et al.'s original study. We tracked participants over the period of one year during which they were exposed to different strengths of natural group identities in their daily lives: one group consisted of first-year students at a large international university (weak group identity) while the others were police cadets in their first year (strong group identity).

Results: We find that discriminatory behavior is very unstable within participants across time and neither robustly associated with other individual level characteristics we measured nor affected by group identity strength. Our results suggest that, if at all, 'groupiness' might be stable only in a very limited sense, as participants who changed their levels of discrimination tended to do so for multiple group markers simultaneously. We are confident that our findings are not rooted in the specifics of our lab-in-the-field setting, as behavior in a neutral dictator game and several psychometric measures that we included in the study show high levels of test-retest reliability even across one year.

Conclusion: Our findings are hard to reconcile with notions of 'groupiness' as a trait. Rather, our results imply that discriminatory behavior is not time-stable. This observation emphasizes, again, the need to understand those situational factors and their interaction with individual characteristics that promote discrimination.

Session 13: Ingroup-Outgroup | Room 1

Thursday July 4, 2024 11:45 - 12:00

**Group-bounded indirect reciprocity and in-group favouritism: recent advancements and future directions**

Hirotaka Imada

Royal Holloway, University of London

Bonded generalized reciprocity (BGR) is one of the major accounts of in-group favouritism in cooperation, the tendency to be more cooperative with in-group members than out-group members. According to BGR, shared group membership cues that indirect reciprocity is bounded by group membership (i.e., the group heuristic). That is, people intuitively assume that in-group members but not out-group members, belong to the same system of indirect reciprocity, resulting in the increased level of expected cooperation (EC) from in-group members and reputational concern (RC) when interacting with in-group members. In this research, I review and discuss two recent advancements in the empirical literature concerning BGR.

1) Bounded, unbounded, and dynamic indirect reciprocity. Despite that BGR has collated empirical and meta-analytical support, some recent studies by Romano and colleagues rather suggest that indirect reciprocity is perceived to be unbounded by group membership. Namely, several experiments revealed that people display RC-based cooperation towards both in-group and out-group members (i.e., unbounded indirect reciprocity, UIR). The literature thus suffered from the mixed evidence with two directly conflicting theories. To reconcile this, the dynamic indirect reciprocity (DIR) perspective was recently proposed with partial evidence from two experiments (Imada et al., 2023). According to DIR, BGR offers a default implicit game strategy for when group membership is a sole cue, but the realm of indirect reciprocity is perceived to include out-group members when cues of future benefits are additionally present; therefore, BGR and UIR are not in conflict but explain intergroup cooperation in different ecologies.

2) Proximate mechanisms underlying in-group favouritism: EC vs. RC.

BGR originally placed EC as the psychological underpinning of in-group favouritism but later incorporated RC as another explanation. However, the relative importance and explanatory power of the two mechanisms have not been discussed until very recently and there is not a clear understanding of how EC and RC shape in-group favouritism. Reviewing recently published papers and some unpublished data, I offer a tentative conclusion; EC, but not RC, seems to explain in-group favouritism but EC is not enough to fully account for it.

Through the review of those two points, I hope to clarify what BGR is and offer the state of the art of the evolutionary and social psychological theory of in-group favouritism.

Session 13: Ingroup-Outgroup | Room 1

Thursday July 4, 2024 12:00 - 12:15

**Humans Exhibit both Parochialism and Nastiness within Groups**

Angelo Romano | Jörg Gross | Carsten K.W. De Dreu

Leiden University | University of Zurich | Leiden University

Decades of research have shown that humans cooperate with ingroup members more than with strangers and individuals from rivaling out-groups. Such parochial cooperation is often taken as suggesting that humans also compete more between than within groups. Whereas this could explain intergroup polarization and conflict, competition is not the flip side of cooperation, and direct evidence for parochial competition is missing. Here we provide a first test of the hypothesis that people compete less with ingroup members than with outgroup members, and unidentified strangers. Against pre-registered predictions, however, people competed systematically and consistently more with ingroup members, than outgroup members and strangers, in a large-scale dyadic contest experiment across 51 nations around the world (N=12,863), in a lab-in-the-field study among tribes in Kenya (N=552), and in a pre-registered replication in the United Kingdom (N=401). Furthermore, people were also more competitive with individuals from geographically (and culturally) closer compared to more distant national outgroups (Study 1 & 3). This 'nasty neighbour' behaviour emerged independent of parochial cooperation and trust towards others that have the same (versus different) nationality (study 4) and, fitting field-observations in other species, neighbour nastiness emerges when people perceive within-group resource scarcity, and especially towards low-ranking ingroup members (study 3-4, and study 5, N=552). That humans can exhibit both parochialism and nastiness within groups is difficult to reconcile with existing theories on the evolution of cooperation in structured populations.

Session 13: Ingroup-Outgroup | Room 1

Thursday July 4, 2024 12:15 - 12:30

**How Group Cooperation Generates Intergroup Conflict**

Laura C. Hoenig | Ruthie Pliskin | Carsten K. W. De Dreu

Social, Economic, and Organisational Psychology, Leiden University | Social, Economic, and Organisational Psychology, Leiden University | Social, Economic, and Organisational Psychology, Leiden University

Per theory and research on the evolution of cooperation, intergroup conflict is a key motivator for individual cooperation within groups. Nonetheless, this scholarship remains silent on how conflict between independently co-existing groups comes about in the first place. Here we fill this void with evidence for a reverse pathway – from cooperation to conflict – showing when and how individual cooperation geared towards maximizing in-group welfare and fairness can – advertently or inadvertently – fuel conflict with other groups.

In two incentivized, no-deception laboratory experiments, individuals organized in two groups repeatedly invested in two club goods, one exclusively benefitting the in-group and one benefitting the in-group at some cost to the out-group. Across decision-rounds we varied the relative efficiency (Experiment 1, N = 132 in 44 groups) and fairness (Experiment 2, N = 210 in 70 groups) of the two club goods, while holding constant the cost to the out-group.

Individuals cooperated more on club goods that were more efficient, even when this created a cost for the out-group, and more on club goods that benefitted in-group members equally rather than unequally, even when this created a cost for the out-group. At the same time, individuals who benefitted from inequality readily imposed costs on the out-group if that created inequality in returns. Costs imposed on out-groups ignited cycles of retaliation and revenge, ultimately reducing welfare at the individual, group, and collective level.

Here, we identify a heretofore unknown reason for intergroup conflict: As much as intergroup conflict can motivate in-group cooperation, it can also be the unfortunate by-product of myopic cooperation within groups. Our findings have implications for theory on the evolution of cooperation and conflict, reveal how cooperatively managing the commons can have tragic consequences for individuals and collectives, and help to understand the origins of conflict within and between human groups and civilizations.

Session 14: Social norms | Room 2

Thursday July 4, 2024 11:00 - 11:15

**Measuring social norms in surveys: The role of question sequence, reference group, and context information in gender norm inquiries.**

Annelie Bruning | Wojtek Przepiorka | Tali Spiegel | Tanja van der Lippe

PhD student | Associate Professor | Associate Professor | Professor

Rationale: Most survey-based studies on the normative underpinnings of individual behavior rely on measures of personal attitudes (i.e. personal normative beliefs) only. This involves the attitudes people have towards statements such as “Men should participate in housework to the same extent as women”. However, recent advancements in the measurement of norms suggest that (1) measuring social norms also requires the elicitation of people’s beliefs about what other people think and do, and that (2) these beliefs exert a greater influence on respondents’ behavior than their attitudes. Many researchers studying norms have already adopted this approach by including social norm inquiries, either before or after the elicitation of attitudes. However, to date there is little evidence on how these inquiries affect each other and the validity of norm measurements.

Methods: To address this research gap, we will conduct a survey experiment on a large and diverse sample of respondents recruited via a Dutch online panel, with data collection set for March, 2024.

Results: First, we examine whether including social norm inquiries before personal attitude inquiries prompts respondents to conform to the social norm they report or to distinguish their personal attitudes from the behavior and beliefs they expect in others (i.e. to report their personal beliefs net of norms). Second, we assess the moderating effect of the closeness of the reference group used in social norm inquiries. We expect the conformity of personal attitudes with perceived norms to be stronger the closer the reference group is. Third, we investigate the effect of context information on attitude questions. This will allow us to examine to what extent responses to attitude questions are distorted by respondents’ assumptions regarding the (non-normative) contextual factors at play.

Conclusions: The results of our study will advance our knowledge on how to best measure social norms in surveys.

Session 14: Social norms | Room 2

Thursday July 4, 2024 11:15 - 11:30

**Avoidance of Altruistic Triggers: Empathy versus Social Pressure**

Ivan Soraperra | Jantsje Mol | Joel van der Weele

Research Scientist at the Center for Human and Machines of the Max Planck Institute for Human Development | Post-Doc at the Microeconomics Section, University of Amsterdam | Professor of Economic Psychology at the Microeconomics Section, University of Ams

Research has established that individuals often avoid opportunities to engage in altruistic behavior, particularly when asked to donate (Andreoni, Rao & Trachtman, 2017). This phenomenon, referred to as "ask avoidance," is suggest to be driven by two mechanisms: anticipated social pressure and/or empathic triggers to give. Individuals who are reluctant to donate may anticipate these triggers and consequently choose to avoid the solicitation altogether.

Our paper investigates both mechanisms in a preregistered online experiment where individuals can decide on whether to donate to Save the Children or not. Our experiment includes two main treatment conditions: an Empathy and a Social Pressure Treatment. In each treatment, participants were asked to state their preference between watching one of two videos before making a donation decision. The “treatment video” was designed to evoke empathy (or social pressure) and the “alternative” video was a neutral video unrelated to charity. Participants were informed of the latter and also told that their preference would be implemented most of the time, but not always. In addition, we included a Control Treatment where both videos were neutral and unrelated to charity.

Data from 1400 UK residents revealed that:

[i.] Pressure and Empathy are effective in increasing donations: Both empathy and pressure videos significantly increased donations by 18% and 10%, respectively.

[ii.]Pressure and Empathy generate avoidance: avoidance levels (preference for the alternative video) were 42%, 58%, and 46% for the Control, Empathy, and Pressure Treatments, respectively. The Empathy Treatment exhibited significantly higher avoidance levels. While avoidance levels of the Pressure and Control Treatment are comparable, open-ended text analysis indicated distinct reasons motivating avoidance in each case: individuals avoided the pressure video mainly to prevent being guilt-tripped to give, while individuals preferred the alternative video in the Control Treatment primarily out of curiosity about the content of the alternative video.

[iii.]Sophistication of the avoidance: The effect of the treatment video on donations was stronger for individuals who did not want to see it compared to the effect on those who did want to see it in the Pressure Treatment but not in the Empathy Treatment.

To conclude, we provide evidence that both empathy and social pressure can trigger ask avoidance, but that the empathic/emotional channel may play a more significant role. Additionally, our findings suggest that individuals do not appear to be as sophisticated in the

anticipation of emotional triggers to give as they are for the anticipation of the effect of social pressure triggers.

Session 14: Social norms | Room 2

Thursday July 4, 2024 11:30 - 11:45

**Risk, sanctions and norm change: the formation and decay of social distancing norms**

Giulia Andrighetto | Eva Vriens | Luca Tummolini

ISTC-CNR; IFFS; LiU | ISTC-CNR; IFFS | ISTC-CNR;

Global challenges like the climate crisis and pandemic outbreaks require collective responses where people quickly adapt to changing circumstances. Social norms are potential solutions, but only if they themselves are flexible enough. The COVID-19 pandemic provided a unique opportunity to study social norm formation and decay in a real-world context. Specifically, tracking norms throughout the pandemic yields insight into the effectiveness of norms in changing contexts (i.e., multiple waves of high/low infection and mortality rates)—particularly when the contextual change may result in changes in the norm itself (i.e. cycles of norm formation and decay).

Methods: We collected data about social expectations of social distancing and sanctioning of people living in Rome (Italy) in two periods: a period of COVID-19 risk decrease (June–August 2021) and a period of risk increase (October 2021–February 2022). Our goal was to study whether norms and meta norms of social distancing covary with risk. During the last two waves, we also presented respondents with hypothetical situations to test whether different norms invoke changes in sanctioning behaviour. Using the strategy method, we manipulated the social norm of distance and meta norm of sanctioning and asked how they would behave in each scenario. We used the scenario of keeping distance in line at the supermarket. While some of the social norms on preventative measures were legally enforced in Italy (e.g. closing restaurants and stores when risk is too high), the social norm of keeping distance from others is purely socially enforced. Governments did declare a minimum distance that should be kept from others in public places (in Italy this was at least 1 m), but had no means to legally enforce this rule.

Results: We found that norms and meta norms partially coevolve with risk dynamics, although they recover with some delay. This implies that norms should be enforced as soon as risk increases. We therefore tested how sanctioning intentions vary for different hypothetical norms and find them to increase with a clear meta norm of sanctioning, yet decrease with a clear social norm of distance.

Conclusions: In conclusion, social norms evolve spontaneously with changing risk, but might not be adaptive enough when the lack of meta norms of sanctioning introduce tolerance for norm violations. Moreover, norm nudges can potentially have negative externalities if strengthening the social norm increases tolerance for norm violations. These results put some limits to social norms as solutions to guide behaviour under risk.

Session 14: Social norms | Room 2

Thursday July 4, 2024 11:45 - 12:00

**The interplay between low- and high-cost cooperation**

Dorothee Mischkowski

University of Leiden, Max Planck Institute for Research on Collective Goods

Costly cooperation behavior in social dilemmas is well understood in its facets and determinants. The concept of low-cost cooperation, known as social mindfulness – the awareness and consideration of others' needs in everyday interactions – has received considerably less attention. In four preregistered, incentivized experiments (N total = 956) the relationship between costly cooperation in economic games and social mindfulness was investigated, including a test of potential moderators.

Next to an expected positive correlation between both forms of cooperation behavior, two interaction effects were hypothesized: First, the relation was expected to increase with opportunity costs of social mindfulness: The more a forgone choice option is valued, not chosen for accommodating another person’s potential preferences, the stronger both forms of cooperation behavior should be related. In a similar vein, the relation was expected to strengthen with decreasing costs of costly cooperation: The lower the stakes of costly cooperation in an economic game, operationalized as monetary endowment in a one-shot public goods game, the stronger the relation between costly cooperation and social mindfulness was expected.

The results consistently revealed a positive association between low- and high-cost cooperation. As hypothesized, this relation was moderated by the opportunity costs of social mindfulness: The more an option is preferred, but not chosen to benefit another person, the stronger the relation to costly cooperation. Contrary to expectations, however, the costs of high-cost cooperation did not moderate the relation between both forms of cooperation behavior.

The findings highlight that the costliness of cooperation is not reflected by the stake size or absolute value of the choice options, but by the subjective cost of self-denial for another's benefit. Future research directions are discussed, probing how low-cost cooperative behavior could catalyze more cost-intensive cooperation in controlled lab environments and everyday life.

Session 14: Social norms | Room 2

Thursday July 4, 2024 12:00 - 12:15

**Personal norms — and not only social norms — shape economic behavior**

Zvonimir Bašić | Eugenio Verrina

Adam Smith Business School, University of Glasgow, 2 Discovery Place, Glasgow G11 6EY, UK | Sciences Po Paris, CNRS, 28 rue des Saints Pères, 75007 Paris, France

We study the relevance of personal norms in economic settings. We propose a simple utility framework, in which people care about their monetary payoff, social norms, and personal norms, and then design a novel two-part experiment to estimate our framework. We show that personal norms — together with social norms and monetary payoff — are highly predictive of individuals' behavior. Moreover, they are: i) distinct from social norms across a series of economic contexts, ii) robust to an exogenous increase in the salience of social norms, which increases the weight people attach to social norms, but does not affect that of personal norms, and iii) complementary to social norms in predicting behavior, as a model with both personal and social norms outperforms a model with only one of the two norms. Our findings show that personal norms are powerful predictors of behavior in economic settings, and they support them as a key motive in economic decisions. Moreover, our results provide evidence for the existence of pitfalls when designing normative interventions or shaping desired behavior within organizations, and they underscore the importance of understanding both personal and social norms in these settings.

**Norm-based and Self-relevant Perceptions of Cooperation**

Shacked Avrashi | Ilan Fischer

University of Haifa, Israel | University of Haifa, Israel

Cooperation is a vague concept - an outcome can be considered cooperative within a certain set of outcomes, but less or even non-cooperative in another. Furthermore, in many cases the perception of cooperation differs when addressing one's own benefits and when addressing those of other individuals. Therefore, the present research tests norm-based (Study 1) and self-relevant (Study 2) attitudes towards cooperation and develops a quantitative model of the perception of cooperative outcomes.

Study 1 tested the perception of cooperation, and its association with related concepts (such as fairness, reciprocity, and coordination). Forty participants were presented with various sets of payoff-pairs, assigned to two fictional individuals. Participants were asked to rate the extent of perceived cooperation (as well as the other related concepts) expressed by the choice of each payoff-pair, in relation to the other payoff-pairs in the presented set. The perception of cooperation was found to be highly correlated with the perceptions of reciprocity (r=.697), coordination (r=.667), and fairness (r=.665). The results also show that equality plays a crucial role in people's reported perception of cooperation, while the joint gain is of less importance. The results allow defining an objective norm-based cooperation index, which assigns a larger weight to payoffs differences than to payoff sums (-10:7 weights ratio).

Study 2 tested the perception of cooperation within a self-relevant setting. Fifty-seven participants were presented with various games, which they played against an anonymous participant. In each game, they were asked to choose one of four payoff-pairs, each pair indicating a payoff for the participant him/herself and a payoff for the opponent. Participants were instructed that they would receive their chosen payoff only if the other participant also chose the same payoff-pair, thus motivating participants to choose what they consider to be the most mutually beneficial outcome. Results show that participants were willing to sacrifice some extent of equality in favor of less equal but larger payoffs for both players. These results suggest a self-relevant cooperation index, which assigns a smaller weight to payoff differences than to payoff sums (-2:3 weights ratio). The novel index correctly predicted 73% of participants' choices.

The two studies reveal a gap between norm-based and self-relevant cooperation attitudes. People report preferring equality when considering general norms; but in practice they sacrifice some equality in exchange for larger joint gains. The derived cooperation indices are expected to help predict strategic behavior in asymmetric interactions.

Session 15: Digital interactions | Room 1

Friday July 5, 2024 09:15 - 09:30

**Time and Ties in Moral Social Dilemmas. How Temporal Distance and Personal Closeness Affect Dishonesty and Moral Judgment**

Yehor Hrymchak | Paul Conway | Katarzyna Cantarero

SWPS University | University of Southampton | SWPS University

According to Construal Level Theory (CLT), considering the future or past evokes abstract processing of information, including reliance on cherished values, compared to concrete processing evoked by considering the present. Reliance on cherished values should, in turn, lead to lower dishonesty for oneself and harsher moral judgments of dishonest others. We examined how temporal delay (now vs. future or now vs. past) influences dishonesty and judgments in social dilemmas across seven studies (N = 2980) and an internal meta-analysis. We further tested whether personal closeness to another person enhances the effect of temporal distance on dishonesty and moral judgments. Overall, results were mixed. The internal meta-analysis showed significant effects of temporal distance manipulations — yet opposite to CLT predictions. Individuals were more prone to dishonestly report in distant temporal settings, compared to present and judged such behavior more leniently, when set further in time. Personal closeness moderated findings: temporal distance impacted dishonesty and moral judgments for close individuals but not strangers. Neither culture, type of dishonesty (prosocial vs egoistic), nor measurement (dishonesty vs. moral judgment) moderated effects. These results suggest that concrete situations like the present reduce dishonesty and increase harsh moral judgments of dishonesty, in line with research suggesting that present settings are more vivid and elicit stronger emotional reactions.

Session 15: Digital interactions | Room 1

Friday July 5, 2024 09:30 - 09:45

**Playing Prisoner Dilemma Games with LLMs**

Andreas Orland | Kazuhiro Takemoto

Corvinus University of Budapest | Kyushu Institute of Technology

Rationale: Artificial intelligence (AI) makes increasingly more decisions for us, including decisions that resemble social dilemmas. We know the general structure of AIs and their training data, but we need to learn more about how they behave or, more precisely, imitate behavior. Understanding this can reveal much about their underlying algorithms and logic, help develop AI that aligns with human values and ethics, help understand how AI might interact with humans, and inform policymakers and regulators overseeing AI technologies.

Methods: We conducted an experiment where ChatGPT, a popular generative AI, made decisions in a one-shot Prisoner's Dilemma. We vary the following dimensions of the underlying game: All payoff parameters (we vary one parameter at a time). The number of simultaneous interaction partners ranges from 1 to 1 million. The strategy space allows for either mixed strategies or is restricted to a pure cooperation/defection decision. We elicit the belief or the decision first, followed by the other variable.

In the one-shot Prisoner's Dilemma, game theory always predicts defection. However, evidence from laboratory experiments strongly suggests that the four dimensions we examine in our study influence human behavior. We use these predictions as a benchmark to compare the AI's behavior imitation. We collected 100,000 observations in 352 experimental treatments (in a full factorial design).

Results: With an overall cooperation rate of 80%, ChatGPT is more cooperative than humans in many lab experiments. Further, ChatGPT reacts to payoff parameter variations (though not always as predicted), decreases cooperation in the number of interaction partners, exhibits a higher rate of cooperation when being restricted to cooperate/defect decisions compared to being allowed to play mixed strategies, and responds with less cooperation when we elicit a belief about the other players' decisions before the decision. Furthermore, decisions and beliefs are more closely related than when the belief is asked only after the decision.

Conclusions: Although ChatGPT is not specifically trained to make human-like decisions in game-theoretic situations, many dimensions resemble human behavior in the one-shot Prisoner's Dilemma. However, it does not imitate behavior perfectly. While ChatGPT is a promising tool for experimentalists (e.g., for power calculations or testing of instructions), it raises concerns as it might also be used for data fabrication.

Session 15: Digital interactions | Room 1

Friday July 5, 2024 09:45 - 10:00

**Ethical Risks of Algorithmic Delegation**

Nils Köbis | Zoe Rahwan | Clara Bersch | Tamer Ajaj | Jean-Francois Bonnefon | Iyad Rahwan

Max Planck Institute for Human Development | Max Planck Institute for Human Development | Max Planck Institute for Human Development | Max Planck Institute for Human Development | Toulouse School of Economics | Max Planck Institute for Human Development

Rationale: A growing exists trend to delegate tasks to algorithms, for example smart pricing options on platforms like Airbnb. Yet, algorithms that autonomously act on people's behalf might break legal or ethical rules, even without humans being aware of it, such as the recent evidence of pricing algorithms colluding autonomously. This raises a fundamental question about Artificial Intelligence safety: can the ability to delegate tasks to machines increase human engagement in unethical behavior?

Methods: To examine this question across different types of algorithmic settings we conducted four pre-registered, large-scale online experiments (total N= 3217). As a measure for ethical behavior we used the die-rolling task, a well-established paradigm where participants are instructed to report the observed outcome of a private die roll but get paid according to the number they report. In all experiments we compare the degree of dishonesty in different delegation settings to a baseline where people report the die rolls themselves.

Results: First, our results show that individuals are more prone to unethical behavior when delegating tasks to machines than when performing these tasks themselves, with only 5-10% behaving unethically in the latter scenario. Second, we show that the manner in which humans program the machines—e.g., using rules, training data, or high-level goals—can qualitatively alter the temptation towards unethical behavior, causing as many as 85% of participants to behave unethically in some conditions. Together, the quantitative increase in opportunities to delegate, coupled with the qualitative change in temptation, combine to substantially increase unethical behavior.

However, we also identified two effective strategies to mitigate this risk: allowing participants to choose whether to delegate to a machine or undertake the task themselves significantly increased honest behavior, with a preference for self-engagement rising to 74% after experience with both human and machine delegation. Additionally, using natural language for delegation, a feature now common in modern chatbot technology, notably reduced unethical delegation.

Conclusion: These results highlight the contexts under which risk of increased unethical behavior arises when delegating to AI and underscore the importance of considering human factors in AI safety.

Session 15: Digital interactions | Room 1

Friday July 5, 2024 10:00 - 10:15

**The Economic Effects of Remote-Bargaining vs. In-Person Bargaining**

Claudia Keser | Holger Rau | Anne Schacht

University of Göttingen, Department of Economics | University of Göttingen, Department of Economics | University of Göttingen, Georg-Elias-Mueller-Institute of Psychology

Our study addresses one of the challenges caused by the increasing reliance on remote work and video telephony in various professional contexts, specifically in negotiations. Our study aims to examine the economic impact of remote negotiations (via Zoom) versus face-to-face negotiations through controlled laboratory experiments. Participants engage in a repeated variant of the Ultimatum Game with communication, negotiating the division of a sum of money. The stake of the negotiation is either low or high and known only to the proposer. Additionally, the study will employ facial-expression analyses to measure the emotions displayed by the negotiators during the two institutional settings.

The results show that bargaining fails less often in remote negotiations, which leads to higher welfare compared to the in-person treatment. While bargaining outcomes tend to be equitable in the majority of cases in the face-to-face negotiations, we observe higher inequity in the remote treatment. This is particularly true in those cases where the stake size is low: the responders often achieve to receive a higher share than the proposer in remote negotiations.

**Corrupted by Algorithms? How AI-generated and Human-written Advice Shape (Dis)honesty**

Margarita Leib | Nils Köbis | Rainer Michael Rilke | Marloes Hagens | Bernd Irlenbusch

Tilburg University | University of Duisburg-Essen | Otto Beisheim School of Management | Erasmus University Rotterdam | University of Cologne

Increasingly, Artificial Intelligence (AI) has become an indispensable advisor, affecting people's behavior. As a case in point, Amazon's chief scientist, envisions that Alexa's role for its over 100 million users "keeps growing from more of an assistant to an advisor". A new ethical concern arises if AI persuades people to break ethical rules. In a large-scale, financially incentivized, pre-registered experiment we examine (i) whether people alter their (dis)honesty following AI-generated advice, and (ii) how such advice compares to human-written advice. Lastly, we test (iii) does knowledge about the source of advice, a common policy recommendation, matters?

First, we recruited participants (N = 367) to write human-written advice, and incentivized them to promote honesty or dishonesty. We then trained a large language model, GPT-J, to generate corresponding AI-advice. Lastly, we recruited another set of participants (N = 1,817), who received advice and then engaged in a die-rolling task in which they could lie for financial profit.

Before the die-rolling task, participants either read an honesty-promoting or dishonesty-promoting advice that was either human-written or AI-generated. Further, participants either knew the source of the advice (transparency) or knew that there was a 50-50 chance that the advice came from either source (opacity). In another control treatment, participants did not receive advice.

Testing human behavior in reaction to actual AI-advice, we find that (i) across all nine treatments participants lie to boost their pay. Further, (ii) in the opacity treatment, where the advice source was unknown, dishonesty-promoting AI-advice increases dishonesty, while honesty-promoting AI-advice does not increase honesty (compared to the no advice treatment), suggesting AI is successful at influencing peoples' (dis)honesty. Moreover, (iii) in the opacity treatment, AI-advice is as persuasive as human-written advice. Examining the common policy recommendation of AI-transparency, we find that (iv) transparency does not alleviate the corruptive influence of AI. That is, even when participants knew the advice was generated by an AI they lied to similar extents as when the advice was written by humans.

Overall, our findings provide first behavioral insights into the corruptive force of AI advice. Whereas previous work has revealed people's stated aversion towards AI making ethical decisions and reluctance to follow AI-advice, our behavioral results show that AI advice corrupts people, even when they know the advice source. Understanding the corrupting power of AI-based advice marks a crucial step towards managing AI responsibly.

Session 16: Reciprocity | Room 2

Friday July 5, 2024 09:15 - 09:30

**Rule Following and Cooperation**

Pierce Gately | Robin Cubitt | Simon Gaechter

University of Nottingham | University of Nottingham | University of Nottingham

Rules and social norms are driving forces of behaviour and constitute the grammar of society by prescribing socially appropriate behaviour in a particular context. We examine how rule following and voluntary cooperation are connected in two preregistered studies (AEA RCT Registry: AEARCTR-0009755). In particular we focus on a non-enforced rule demanding cooperation. Study 1 documents differences in behaviour, while Study 2 examines differences in social expectations and conditional preferences.

The process of rule following is governed by both personal and social motives. In these experiments we examine both motives side-by-side with participants either facing a rule that has a social purpose (i.e., the provision of a public good) or a rule that has no purpose (i.e., standard Rule-Following Tasks (RFT)). Combining features from Public Goods Games (PGG) and RFT allows us to analyse the relationship between rule following and cooperation.

In Study 1 we examine the impact of a rule on cooperative behaviour, and differences in rule following behaviour across three classifications of rules: (i) pointless rule; (ii) pointful rule; and (iii) pointful rule with stated purpose. In online experiments (N = 408) we highlight the important role expectations of others behaviour play in driving one's own behaviour and replicate previous findings that introducing a rule into PGG frameworks increases the rate of cooperation.

In Study 2 we analyse underlying mechanisms which may explain the differences in behaviour observed in Study 1. We examine differences in social expectations (normative and empirical expectations) and conditional preferences (both normative and empirical) in the tasks participants completed in Study 1. That is, in Study 2 we analyse differences in social expectations and conditional preferences for three classifications of rules: (i) pointless rule; (ii) pointful rule; and (iii) pointful rule with stated purpose. Data collection for Study 2 (N = 3,108) is ongoing, wherein we examine how the consequences of a rule alter conditional preferences and social expectations.

Session 16: Reciprocity | Room 2

Friday July 5, 2024 09:30 - 09:45

**Young children protect rule-breakers whom they owe a favor**

Sebastian Grüneisen | Tindaya Déniz | Louisa Huff

Leipzig University | Leipzig University | Leipzig University

Rationale: Sanctioning norm violations is critical for maintaining cooperation. Yet, norm violations often remain selectively unpunished depending on the norm enforcer's relationship with the transgressor. Examples include gender-based double standards (e.g., Lai & Hynie, 2010), ethnicity-based discrimination in the courtroom (Gazal-Ayal & Sulitzeanu-Kenan, 2010; Shayo & Zussman, 2011), and the well-documented informal code of silence among police officers not to report on a colleague's misconduct (Skolnick, 2002). In these cases, the selective sympathy and obligation the enforcer experiences toward the transgressor directly clash with the principle of impartiality.

Young children show an early-emerging tendency to reciprocate favors (House et al., 2013; Vaish et al., 2018) and view doing so as obligatory (Wörle & Paulus, 2019). Yet, children also show strong motivations to sanction norm violations (Marshall & McAuliffe 2022; Rakoczy & Schmidt, 2013). A largely unexplored question is whether the sense of obligation emanating from reciprocal cooperation can compromise children's tendency to enforce norms equally.

Methods & Results: In a series of preregistered studies, we investigated children's evaluation of and participation in unequal norm enforcement. Studies 1a and 1b (n = 48 and n = 72, respectively) demonstrated that children aged 5 and older disapprove of individuals who enforce norms unequally, but accept both unequal treatment when justified (Study 1a) and leniency (i.e., no enforcement) when applied consistently (Study 1b). Study 2 (n = 72) showed that, from age 6-7, children accept mitigating circumstances as a reason for unequal norm enforcement but condemn enforcers who selectively spare transgressors with whom they have a special relationship (e.g., nepotism).

Study 3 consisted of a behavioral experiment investigating if children engage in unequal norm enforcement themselves. 5- to 7-year-olds (n = 85) witnessed a game partner cheat. This partner had either previously done them a favor (reciprocity condition) or behaved neutrally (control condition). Across ages, children were slower and less likely to tattle on the transgressor to the experimenter in the reciprocity than the control condition, both spontaneously and when asked directly, necessitating a lie. In contrast, the same children often advised others to tattle on transgressors, revealing an intriguing knowledge-behavior gap.

Conclusions: The studies demonstrate that young children are sometimes willing to flout rules they otherwise approve of in favor of reciprocating prosocial acts. They also showcase a "dark side" of reciprocal obligation: From a young age, children protect transgressors who helped them in the past, even though they know better.

Session 16: Reciprocity | Room 2

Friday July 5, 2024 09:45 - 10:00

**Cooperating Across Generations: Experimental Evidence of Reciprocal Cooperation and Intergenerational Exchange**

Ben Grodeck | Zach Freitas-Groff | Oliver Hauser | Johannes Lohse

Max Planck Institute for Evolutionary Biology | Stanford University | University of Exeter | University of Birmingham

Within both families and states, important policies involve pay-it-forward schemes where one cohort invests in the next cohort (e.g., via education), sometimes with an expectation of a future reward (e.g., retirement support). Because of the upfront cost of such pay-it-forward schemes, it is unclear to what degree they can be sustained through private contributions. We study altruism, reciprocity, social information, and self-interested equilibria as possible mechanisms to sustain intergenerational cooperation. We conduct a large-scale online experiment where a sequence of players choose how to allocate an endowment between themselves and future, prior, or contemporary players. By varying the option set and the information set, we can separate possible mechanisms for sustaining investment in the future. We find that the ability to give back increases the willingness to give forward, but a social preference for reciprocity rather than individual rationality drives the result.

Session 16: Reciprocity | Room 2

Friday July 5, 2024 10:00 - 10:15

**Proactive Cooperation and Reciprocation in Real-time PD**

Ryutaro Mori | Nobuyuki Hanaki | Tatsuya Kameda

The University of Tokyo | Osaka University | The University of Tokyo

Rationale: The prevalence of conditional cooperators—individuals who are willing to cooperate if others do—may suffice for overcoming the free-riding problem and launching a collective action once the "initial level" of cooperation is provided. However, it remains unclear how this initial level of cooperation is established, either through observable cooperative acts or optimistic expectations in agents' minds. Previous studies, predominantly assuming "turn-based" interactions where agents choose actions simultaneously or sequentially, have limited agents to passive roles in initiating cooperation; They either rely on uninformed expectations or just follow exogenously assigned decision orders. By contrast, many real-world interactions unfold in real time, allowing agents to make and communicate their decisions instantly. We propose that in such a setting, some agents may proactively choose to cooperate first, hoping to trigger reciprocal actions from conditional cooperators. The fact that agents can choose their decision timings in real-time interactions can thus facilitate collective actions, even with the risk of free riding.

Methods: We extend the Prisoner's Dilemma game to a real-time protocol. Specifically, pairs of participants have 60 seconds to make their decisions at any time. These decisions are final and immediately communicated within the pair. If players do not decide during the 60 seconds, they will be asked to decide at the end (if neither player in a pair decides, they will decide simultaneously). We contrasted real-time interactions with simultaneous (Experiment 1) and sequential turn-based (Experiment 2; to be conducted) interactions in a between-subject design. In Experiment 1, we recruited 152 participants using Prolific Academic, with all hypotheses and procedures, including sample size, preregistered.

Results: The results from Experiment 1 showed that in the simultaneous condition, 71.1% of participants chose cooperation, resulting in less than half of the pairs (47.4%) reaching mutual cooperation. In the real-time condition, all pairs decided before the 60 seconds had passed, with 89.5% of first movers in a pair choosing cooperation. When these first movers cooperated, 94.1% of second movers returned cooperation, leading to mutual cooperation in 84.2% of pairs. Moreover, the first movers within a pair were significantly more cooperative and slower (KS-test; $D = 0.42$; $p < 0.01$) in the real-time condition, suggesting the strategic intentions of the first movers.

Conclusions: We have theoretically explored the differences between simultaneous or sequential turn-based interactions and real-time interactions. Our experiment highlights how real-time interactions, a previously under-investigated structural factor, assist pairs in overcoming the temptation to free ride.

Session 16: Reciprocity | Room 2

Friday July 5, 2024 10:15 - 10:30

**Cooperation and strategy choice in the infinitely repeated prisoner's dilemma when players can cheat**

Astrid Dannenberg | Marcel Lumkowsky

University of Kassel, Germany | University of Kassel, Germany

We study the impact of imperfect monitoring structures in an infinitely repeated prisoner's dilemma when players can cheat by signaling cooperation while actually playing defection. Players in this novel extension of the game decide whether to cheat or play honestly while nature decides if a cheating attempt succeeds or fails. Previous studies on imperfect monitoring structures focused on the impact of inadvertent, exogenously induced errors, and thus, overlooked scenarios where false signals are induced deliberately. Nevertheless, intentional errors are common in daily life, ranging from infidelity in relationships to fraudulent activities in corporate settings or the use of performance-enhancing substances in sports.

In a laboratory experiment, we test how the possibility to cheat affects strategy choice and cooperation compared to the classic version with perfect monitoring where players are unable to cheat. We also test how the probability that a cheating attempt succeeds or fails and whether the co-player's ability to discover failed cheating attempts change behavior. In total, 356 undergraduate students were randomly allocated into one of five experimental conditions (between-subjects design) and made at least 108 decisions over the course of a session (with different co-players). With every decision incentivized, subjects earned 23€ on average.

Although our parametrization theoretically allows for cooperative equilibria in all treatments, cooperation rates are significantly reduced (based on two-way clustered probit regressions) when cheating is possible. The negative effect of cheating on cooperation rates is stronger when a failed cheating attempt is not revealed as such to the co-player. Notably, the probability of success seems to have little effect on the use of the cheat option. By applying the "Strategy Frequency Estimation Method (SFEM)", we gain further insights by categorizing subjects' behavior into underlying game theoretic strategies. We demonstrate that the decline in cooperation rates can be attributed to a large proportion of players opting for unconditional defection when cheating is an option.

To the best of our knowledge, our experiment is the first in the context of the infinitely repeated prisoner's dilemma where players have an influence on whether the monitoring structure is imperfect or not. The findings have implications for real-world applications where monitoring structures allow for self-reporting with imperfect verification. The results suggest that the possibility of cheating poses a significant threat to successful cooperation in these situations even when chances of successful cheating are relatively low and cooperative equilibria are theoretically possible.

**The repeated punishment game explains why, and when, we seek revenge**

Julien Lie-Panis | Bethany Burum | Christian Hilbe | Moshe Hoffman

Max Planck Institute for Evolutionary Biology | Harvard University | Max Planck Institute for Evolutionary Biology | Harvard University

Rationale: A prominent explanation for revenge is the deterrence of future transgressions. Yet revenge often fails to achieve optimal deterrence. We forgive others' dangerous behavior based on variables out of their control, such as a lucky positive outcome. Minor offenses can lead to full-blown conflict, rather than a proportionate response. In addition, the role of apologies remains unclear---when do apologies need to involve costs, and when are mere words enough?

Methods: Here, we address these gaps in our knowledge of revenge using a mathematical model. This model—the repeated punishment game—involves one actor and many successive partners. Each partner decides whether or not to transgress on the actor. The actor then decides whether or not to punish a partner.

Results: To guarantee partner cooperation (i.e. non-transgression) in an evolutionary equilibrium, we show that the actor must punish transgressors, and that such retaliatory punishment—revenge—then serves to deter future partners from transgressing.

We show that apologies should similarly serve a deterrence function, in order to maintain a cooperative equilibrium. When transgressions benefit partners, they should be costly; in contrast, they can be cost-free when transgressions do not benefit partners (e.g., because they are accidental).

Finally, we show that revenge can be based on categorical distinctions (e.g., whether a transgression occurred), but not on continuous variables (e.g., the magnitude of the transgression).

Conclusion: Our model helps us better understand revenge and forgiveness. It makes specific predictions about the effects of revenge, and the effectiveness of (costly) apologies.

In addition, our model shows that it's possible that revenge evolved to deter harm, even if it does so in a seemingly sub-optimal manner. Many of the apparent quirks of revenge can be explained in terms of discounting information that occurs on a spectrum (e.g., the riskiness of an offender's behavior, the magnitude of the offense, or the likelihood of observation) in favor of information that is split into categories (e.g., the outcome of a risky behavior, or the type of offense that was committed). In contrast to what other researchers have advanced, these quirks do not disprove the deterrence account of revenge—as our model demonstrates, both deterrence and these quirks can emerge from an evolutionary process.

Session 17: Punishment & Reward | Room 1

Friday July 5, 2024 11:15 - 11:30

**Monetary sanctions are more effective than (dis)approval in maintaining long-term cooperation**

Pat Barclay | Adam Sparks

University of Guelph | University of Guelph

People contribute more to public goods when they can be punished for not doing so, but punishment is costly for both punisher and recipient. As such, some research suggests that it might be easier to maintain public goods using non-costly punishment, such as verbal punishment or disapproval, given that most people avoid doing things that bring disapproval. However, we propose that because humans are avid learners, they will eventually habituate to punishment that carries no tangible consequences (e.g., non-monetary punishment like disapproval). To test this, we present two experiments using 4-person 40-round public games with different types of feedback: monetary sanctions, cost-free (dis)approval ratings, and no feedback (i.e., a standard public goods game). In Experiment 1 (N=224), disapproval and monetary punishment elicited similar levels of public goods contributions in the early rounds, but disapproving feedback eventually ceased to become effective, and was eventually no better than no feedback. In Experiment 2 (N=424), we added an emoticons condition, and used both positive and negative sanctions (i.e., monetary rewards + punishment vs. approval + disapproval ratings vs. smiling + frowning emoticons vs. no feedback). Contributions were initially similar under all three types of feedback, but contributions continued to rise when there were monetary sanctions but stopped rising when there were only (dis)approval ratings and emoticons. Together, these results show that rewards and punishment are most effective in the long run if they have tangible consequences, like monetary bonuses an penalties, as disapproval-avoidance alone is insufficient to maintain long-term cooperation. In other words, because people habituate to non-consequential punishment, punishment must have “teeth”.

Session 17: Punishment & Reward | Room 1

Friday July 5, 2024 11:30 - 11:45

**Public feuds and cooperation**

Yamit Asulin | Yuval Heller | Hillay Zmora | Ro'i Zultan

Ben-Gurion University of the Negev and Bar-Ilan University | Bar-Ilan University | Ben-Gurion University of the Negev | Ben-Gurion University of the Negev

The Feud institution that emerged in early modern Germany poses a puzzle to historians: “Nobles tended to feud against the very people from whose goodwill they had much to gain and from whose enmity much to lose” (Zmora, 2020). We build on the explanation put forward in Zmora (2020) to propose that feud institutions solve the inherent difficulties in community enforcement. The literature on the emergence of cooperation stresses the role of indirect reciprocity: to support cooperation, individuals punish those who transgressed against others. For this mechanism to work, individuals must know the complete history of the whole society to be able to distinguish between acts of transgression and acts of punishment. We suggest that the feud institution relieves this implausibly strict epistemic requirement. In our model, individuals in the society play pairwise Prisoner’s Dilemma (PD) games. After the interaction, each individual can announce a feud on her partner. A feud is similar to standard punishment in the PD/Public Goods Game literature, with the addition that other players in the group can observe a public feud. We characterize an equilibrium where indivduals do not cooperate with those who were punished in the previous period. Thus, the role of a feud is not to directly punish the transgressor, but to announce the trangression to the society. We show that for a range of parameters for which cooperation is unstable without feuds or with private feuds, public feuds can support cooperation in equilibrium. We test the efficacy of public feuds in sustaining cooperation in a laboratory experiment. In Each period, each player play a two-person PD game with each of the other players in a group of six. Players cannot identify others from round to round. This design choice effectively eliminates direct reciprocity, allowing us to identify the role of indirect reciprocity and public feuds. We compare three treatments: a baseline with no punishment; a private feuds treatment, in which a round of punishment follows the PD game in each period; and a public feuds treatment. Public feuds are similar to private feuds, with the addition that at the time of making the choice whether to cooperate or defect, a player knows for each partner how many of the other players in the group (up to four) punished the partner in the previous period. The results confirm the theoretical analysis at the individual and at the institutional level.

**Luck that builds merit**

David Munguia Gomez | Emma Levine

Yale School of Management | University of Chicago Booth School of Business

People find it unfair to reward candidates based on luck but fair based on merit. However, luck can lead to the development of merit. For example, being born into wealth can provide access to better education and training, thus developing skill. For those who have to decide how to allocate rewards, this intersection of luck and merit creates a dilemma between the unfairness of luck-based allocation and the fairness of merit-based allocation. Our research explores how people allocate rewards, such as college admission and employment, and assess fairness when evaluating candidates who 1) benefitted from luck that contributed to developing skill, 2) simply benefited from luck (without developing skill), or 3) are skilled without apparent influence of luck.

In a within-subjects experiment, 151 Prolific participants evaluated 10 Players for a high-paying Wordle tournament, yielding 1,510 evaluations. Participants rated the fairness of selecting each Player and their likelihood of inviting them to the tournament. Players' profiles were based on a previous experiment where 61 participants played Wordle puzzles under different luck-based situations. Player profiles had identical performance but included different information about their effort and previous luck-based situation, based on our experimental design: 2-Source of Luck Benefit (External/Internalized) x 2-Impact (Limited/Lasting) x 2-Effort Required (Low/High), along with a comparison of Just Merit (no luck) where the Effort Required was Low/High. Our analyses here focus on three Players exemplifying luck that builds skill (Internal, Lasting, High Effort), pure luck (External, Limited, Low Effort), and pure skill (Merit, High Effort).

We observed a disparity between perceived fairness and reward decisions. Participants found it less fair to reward the candidate who was lucky to build skill than the candidate with skill not seemingly influenced by luck (difference between means = .63, paired t-test: $t[150] = 5.81$, $p < .001$), but fairer than the candidate who was simply lucky (difference = .62, paired t-test: $t[150] = 4.86$, $p < .001$). Despite this, participants rewarded the candidate with luck-developed skill as much as the candidate with just skill (difference = .01, paired t-test: $t[150] = 0.09$, $p = .939$).

These results highlight people's nuanced views on luck and achievement. People not only consider the presence of luck in achievements, but also the way luck influenced those achievements. This helps understand why people may view certain inequalities as unfair, like the correlation between wealth and college admissions, yet may reinforce them through their decisions.

**Dimensions of Transgressive Social Interaction: A conjoint experiment**

Jolien van Breen | Jarek Kantorowicz | Marieke Liem

Leiden University | Leiden University | Leiden University

Social dilemmas often arise in the context of strained social relationships or transgressive behaviour more generally. Transgressive behaviours are those that transgress social norms regarding interpersonal behaviour. The concept of transgressive behaviour is key to various literatures, not only within (Social) Psychology – punishment decisions in social dilemma games are often based on perceptions of transgressive behaviour - but also in Criminology and Sociology. However, little is known about how individuals determine the (un)acceptability of a certain behaviour – how they "draw the line" between acceptable and unacceptable behaviour. In this work, therefore, we study the decision criteria people use to determine the (un)acceptability of various kinds of negative social behaviours. Through the LISS panel, we recruited a sample of 724 participants who constitute a representative cross-section of the Dutch population. We used a conjoint design to present participants with scenarios describing negative social interactions, ranging from physical attacks, to shouted insults. We experimentally manipulated 6 features of the scenarios, namely the Tactic used by perpetrator; Injury done to victim; Motivation of Perpetrator; Relationship between victim and perpetrator; Gender of the Victim, Group membership of victim. We then asked participants whether they found the behaviour described in the scenarios transgressive. We observed that all 6 manipulated elements affected ratings of transgression. The patterns of differences suggested that participants use 3 criteria in determining the (un)acceptability of a behaviour. First, the adverse impact of the event on the victim was a particularly strong predictor of the transgressive nature of an event – scenarios where a victim is physically injured were rated as particularly transgressive. Further, negative interactions for which there was no obvious explanation (e.g. an attack by a stranger) were rated as more transgressive than those with an implied explanation (e.g. poor mental health of perpetrator). Finally, the implied injustice of an interaction is key in the unacceptability of the behaviour. Participants found it highly transgressive when a perpetrator targets a victim who can be seen as "weak", such as when the victim is elderly or disabled. Similar effects were observed for victim-offender power differentials that were not based on physical ability but on social inequality. Participants found scenarios where members of disadvantaged groups were targeted more transgressive than an equivalent scenario where the background of the victim is unknown. We argue these findings are relevant to understanding participants' responses to social dilemma's.

**Game theory of the hometown tax system**

Ivan Romic

Center for Computational Social Science, Kobe University

Public goods, ranging from clean environment to judiciary to public infrastructure, play a crucial role in societal welfare and development. Yet, these essential services often suffer from underfunding, especially in the aging nations and their depopulated rural areas. To counteract this, Japan introduced the hometown tax (furusato nozei) system in 2008, allowing taxpayers to allocate a portion of their resident tax to municipalities other than their own. This innovative tax policy aimed to invigorate competition among municipalities and foster the redistribution of wealth from affluent urban centers to disadvantaged rural locales. Since its inception, the hometown tax system has undergone several iterations, with a significant increase in donations post-2015, attributed to the introduction of thank-you gifts as part of the donation process. However, this incentive led to heightened competition among municipalities through gift offerings rather than through enhancements in government efficiency and public goods development, necessitating additional adjustments to the system.

Despite its practical evolution, academic scrutiny of the system remains sparse, often limited to analyses from the perspectives of tax and public choice theories. Yet, the system provides a fertile ground for broader inquiry, including the development of public and common goods and insights into human behavioral patterns, as taxpayers' choices reflect a social dilemma, i.e. a balance between altruistic motives (donating to underdeveloped or disaster struck municipalities) and self-interest (donating to municipalities with best gifts).

This paper encapsulates the hometown tax system into a public goods game. In the classical public goods game, participants decide whether to contribute endowment to a communal pot that symbolizes a public good. Contributions are multiplied by a predetermined factor and then equally distributed among all group members. This setup allows for the possibility of free riding, as individuals who contribute nothing can still benefit from the multiplied contributions of others.

Here, we develop a model where individuals have the freedom to allocate their endowments across multiple public goods, paralleling the choices available in the hometown tax system. The model examines the competitive dynamics and variation between public goods, as well as development of these goods through sustained contributions, demonstrating the applicability of game theory in refining the hometown tax system.

**Indirect reciprocity with stochastic and dual reputation updates**

Yohsuke Murase | Christian Hilbe

RIKEN R-CCS | Max Planck Institute for Evolutionary Biology

Cooperation is a crucial aspect of social life, yet understanding the nature of cooperation and how it can be promoted is an ongoing challenge. One mechanism for cooperation is indirect reciprocity. According to this mechanism, individuals cooperate to maintain a good reputation. This idea is embodied in a set of social norms called the “leading eight”. When all information is publicly available, these norms have two major properties. Populations that employ these norms are fully cooperative and stable against invasion by alternative norms. In this study[1], we extend the theoretical framework of the leading eight in two directions. First, we include norms with ‘dual’ reputation updates. These norms not only assign new reputations to an acting donor; they also allow for updating of the reputation of the passive recipient. Second, we allow social norms to be stochastic. Such norms allow individuals to evaluate others with certain probabilities. Using this framework, we analytically characterize all evolutionarily stable norms that lead to full cooperation in the public information regime. When only the donor’s reputation is updated, and all updates are deterministic, we recover the conventional model. In that case, we find two classes of stable norms: the leading eight and the ‘secondary sixteen’. Stochasticity can further help to stabilize cooperation when the benefit of cooperation is comparably small. Moreover, updating the recipients’ reputations can help populations to recover more quickly from errors. Overall, our study highlights a remarkable trade-off between the evolutionary stability of a norm and its robustness to errors. Norms that correct errors quickly require higher benefits of cooperation to be stable.

[1] Y.Murase & C.Hilbe “Indirect reciprocity with stochastic and dual reputation updates" PLOS Comp. Biol. (2023)

Session 18: Games & Game Theory | Room 2

Friday July 5, 2024 11:30 - 11:45

**The evolution of boundedly rational learning in games**

Marta C. Couto | Fernando P. Santos | Christian Hilbe

Max Planck Institute for Evolutionary Biology | University of Amsterdam | Max Planck Institute for Evolutionary Biology

Rationale: Social behavior is often modeled using the mathematical framework of game theory. In its classical form, game theory presumes that individuals are perfectly rational: they know the incentive structure of the game and can always compute optimal strategies. Differently, in evolutionary game theory, players do not need to adopt optimal strategies from the outset. However, there is a usually fixed parameter – selection strength – that regulates how often individuals select strategies that might provide higher payoffs. In this sense, we can interpret selection strength as sensitivity to payoffs or as rationality. While rationality has an important role in strategic decision-making, its origins are hardly discussed.

Methods: If higher rationality allows individuals to better discriminate high-payoff strategies, one would expect that it has an evolutionary advantage. However, in situations where the collective and personal optimal outcomes do not coincide, it is not clear whether full rationality will always evolve. As such, we investigate the evolution of rationality for several social dilemmas. We assume individuals interact with each other and learn their strategies by repeatedly assessing payoffs. Importantly, each individual can follow a more or less rational learning rule. At a much slower time scale than the learning process, rationality evolves. Here, we use the adaptive dynamics framework, assuming rationality is a continuous trait.

Results: We show that the evolutionary endpoint depends on the game. For most prisoner's dilemmas, rationality evolves towards an ever-increasing value. Surprisingly, however, we find that in prisoner's dilemmas for which the temptation to defect is small compared to the benefit of mutual cooperation, the evolutionary endpoint is a finite value. We observe a similar result for the majority of the snowdrift games. Differently, for some stag-hunt games, we find evolutionary branching. This means that, eventually, the population will be composed of individuals with different levels of rationality.

Conclusions: We show that full rationality does not always evolve. Notably, this result does not depend on the assumption that higher rationality requires higher cognitive costs. In some scenarios, a seemingly erratic behavior can, in fact, be strategically justified. These results resonate with earlier ideas and observations in economics and psychology that humans have evolved bounded rationality. Also, our model sheds light on how evolution shapes learning mechanisms.

Session 18: Games & Game Theory | Room 2

Friday July 5, 2024 11:45 - 12:00

**Reactive strategies with longer memory**

Nikoleta E. Glynatsi | Christian Hilbe | Martin Nowak

Max Planck Institute for Evolutionary Biology | Max Planck Institute for Evolutionary Biology | Harvard University

Objective: Repeated games have been the centerpiece in studying the evolution of human cooperation. In these games, strategies represent the strategic rules employed by players when deciding on an action in each turn, given past interactions. Traditional theoretical models assume that players' strategies are limited to remembering only the preceding turn. This limitation arises because, as players remember more rounds, the dimension of the strategy space increases exponentially, making it challenging to derive analytical results.

Here, we explore strategies that remember beyond the last round by concentrating on a specific set of strategies that react solely to the co-player’s actions in the previous n rounds—these are referred to as reactive-n strategies.

Methods &Results: We developed an algorithm to verify whether a given reactive-n strategy is a partner. Partner strategies are Nash strategies that ensure mutual cooperation. We mathematically characterize all partner strategies among reactive-2 and reactive-3 strategies. Additionally, we examine a specific subset called reactive counting strategies, which only consider the count of cooperations in the last n turns. For counting strategies, we characterize partner strategies for all memory lengths.

We establish this series of results by relying on a central finding: if a player employs a reactive strategy, then the co-player using a memory-n strategy can switch to a self-reactive-n strategy without altering the resulting payoffs.

Furthermore, we evaluate the evolutionary properties of partner strategies in the context of higher memory. Our findings suggest that partner reactive strategies evolve more prominently with increased memory, despite the greater number of available strategies. As memory increases and more partner strategies evolve, cooperation rates within the population also increase.

Conclusion: Our work represents a significant advancement in analyzing higher-memory strategies for repeated games. Our evolutionary analysis shows that more memory is beneficial in an evolutionary context. Contrary to the argument in the literature that in a one-to-one interaction, longer memory yields diminishing returns, we contribute to the conversation by highlighting that, in an evolutionary context, individuals who remember more can achieve more.

Session 18: Games & Game Theory | Room 2

Friday July 5, 2024 12:00 - 12:15

**Cooperative dilemmas with binary actions and multiple players**

Jorge Peña | Georg Nöldeke

Department of Social and Behavioral Sciences, Toulouse School of Economics | Faculty of Business and Economics, University of Basel

The prisoner's dilemma, the snowdrift game, and the stag hunt are two-player symmetric games that are often considered as prototypical examples of cooperative dilemmas across disciplines. However, surprisingly little consensus exists about the precise mathematical meaning of the words “cooperation” and “cooperative dilemma” for these and other binary-action symmetric games, in particular when considering interactions among more than two players. Here, we propose definitions of these terms and explore their evolutionary consequences on the equilibrium structure of cooperative dilemmas in relation to social optimality. We show that our definition of cooperative dilemma encompasses a large class of collective action games often discussed in the literature, including congestion games, games with participation synergies, and public goods games. One of our main results is that regardless of the number of players, all cooperative dilemmas—including multi-player generalizations of the prisoner's dilemma, the snowdrift game, and the stag hunt—feature inefficient equilibria where cooperation is underprovided but cannot have equilibria in which cooperation is overprovided. We also find simple conditions for full cooperation to be socially optimal in a cooperative dilemma. Our framework and results unify, simplify, and extend previous work on the structure and properties of cooperative dilemmas with binary actions and two or more players.

Session 18: Games & Game Theory | Room 2

Friday July 5, 2024 12:15 - 12:30

**When games get worse**

Bryan Bruns

Independent Researcher and Consultant

Rationale: How can we understand the ways in which changing social situations may get worse, posing more challenges for cooperation? Changes in the structure of outcomes and incentives may make a situation better or worse, make it easier to find a good solution or more difficult. The solution to a social dilemma is often to realign incentives so that cooperation is a stable, win-win outcome, transforming the situation into a better game, such as turning a Prisoner's Dilemma into a Stag Hunt. However, changes could also go in the opposite direction, moving away from win-win situations, for example, if monitoring becomes difficult, sanctions are not enforced, reputation stops mattering, or selfish gains become too tempting. Shifts in payoffs, the relative ranking of outcomes, and the resulting incentive structures could increase conflict, make it harder to find solutions, and lead to poorer outcomes.

Methods: Simple two-person two-choice (2x2) game theory models of interdependence in strategic situations, such as Prisoner's Dilemma, Stag Hunt, and Chicken, have played a central role in social theory and research on cooperation. This paper applies the topology of 2x2 games, based on changes in the ranking of outcomes (payoff swaps), to explore vulnerabilities and risks in social situations, showing potential pathways to disadvantage, instability, coordination problems, and social dilemmas. Visualizing the topology of 2x2 games as a periodic table of interdependence offers a map for understanding and analyzing the payoff space of 2x2 situations, including risks and vulnerabilities in changing situations. Retelling stories associated with classic 2x2 games illustrates the potential for adverse changes.

Results: Examining potential changes in 2x2 strategic situations shows how stag hunts could turn tragic, concord fall into fighting, coordination become a battle, and harmony decay into disadvantage. Situations differ in their resilience or vulnerability to changes in the ranking of outcomes. A map of potential transformations and new versions of old stories provide tools for thinking about how people might act in the context of changing payoffs, and a framework for considering theoretical and empirical ways to study the diversity of dynamic and uncertain situations.

Conclusions: The topology of 2x2 games and periodic table of interdependence provide a systematic way to display, understand, and examine adverse changes in social situations and how this might affect cooperation and conflict.

Short session 1 | Room 1

Tuesday July 2, 2024 09:00 - 09:05

**Can we measure trait morality? Revisiting the Moral Character Questionnaire and introducing a new, integrative self-report scale**

Nicole Casali | Alicia Seidl | Isabel Thielmann

Department of Criminology, Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany | Department of Criminology, Max Planck Institute for the Study of Crime, Security and Law, Freiburg, Germany | Department of Criminology, Max Plan

Studying and measuring morality has gained considerable traction in contemporary psychology and the broader societal discourse. As the demand for measures of individual differences in trait morality – that is, the way in which people think, feel, and behave in line with ethical principles – grows, recent efforts have attempted to address this need and, for example, proposed the Moral Character Questionnaire (MCQ; Furr et al., 2022). In Study 1 of this project, we critically evaluated the psychometric properties of the MCQ in a large (N = 7,754) and demographically diverse, German-speaking sample that completed the MCQ alongside a variety of other measures of self-reported personality traits and moral behaviors (i.e., dishonesty and prosociality). The results showed model fit issues, alongside poor convergent and predictive validity, thus severely questioning the usefulness of the MCQ. These findings prompted us to develop a new instrument to assess trait morality in a reliable and valid way. First, based on an extensive literature search, we identified five core dimensions of morality, namely honesty, humility, fairness, compassion, and gratitude. We then selected items from existing self-report measures and generated new items designed to capture typical situations and behaviors related to the five morality dimensions to create a new, integrative measure, the Trait Morality Questionnaire (TMQ). Using ant colony optimization, in Study 2 (N = 550 English-speaking Prolific participants) we selected a refined set of 50 items (10 per dimensions) based on model fit, item key balance, and minimum loading criteria. We are now testing this final item set in another study in which we are going to examine its factorial structure, convergent validity, and measurement invariance across two language versions (English and German). We have also created a short version (25 items, 5 per dimension) that is currently being tested, including convergent and divergent validity measures and test-retest reliability. Based on the results of these two additional studies, in Study 5 we will analyze the TMQ's convergence between self- and observer-reports and further extend its nomological network. All these results will be presented at ICSD. Overall, we are confident that our new measure not only addresses the identified shortcomings of existing instruments but also offers a refined and multifaceted perspective on the construct of trait morality. As such, this project contributes to the ongoing dialogue by providing a robust tool for researchers and practitioners to validly assess individual differences in dispositional morality.

Short session 1 | Room 1

Tuesday July 2, 2024 09:05 - 09:10

**Power imbalance leads to out-group aggression**

Nobuhiro Mifune | Hirotaka Imada | Yoshio Kamijo | Atsushi Tago

Kochi University of Technology | Royal Holloway, University of London | Waseda University | Waseda University

Intergroup conflict is ubiquitous in human societies. However, previous studies consistently reported that people do not display increased out-group aggression. Thus, there remains a puzzle; What triggers out-group aggression? We focus on the power imbalance of aggression.

In this study, we conducted two laboratory experiments using a preemptive strike game to investigate the influence of power imbalance of out-group aggression in minimal groups. In this game, a pair of participants were given 800 yen from the experimenter, and they decided whether or not to press the button displayed on a PC screen within a time limit. If neither of them pressed the button, they both ended up with the initial endowment (800 yen). Pressing the button cost 100 yen but reduced the opponent's payoff by 400 or 800 yen. If both participants pressed the button, the first one to do so won, and the loser was not able to attack the winner. Participants played this game against a minimal in- or out-group member.

There were three conditions in the two experiments: equal, weak, and strong conditions. In the equal condition, both participants could reduce the payoff of their opponent by 400 yen. In the weak condition, they could reduce the payoff of the opponent by 400 yen, but the opponent could by 800 yen. In the strong condition, they could reduce the payoff of the opponent by 800 yen, but the opponent could by 400 yen.

The result of a pretest with 222 participants showed that there was no difference in the likelihood of attacking between the in- and out-group members in equal and strong conditions. Yet, they attacked the out-group member more than the in-group member in the weak condition.

Because the pretest lacks comprehensive checks of the manipulation, we conducted a preregistered main study. Results with 347 participants showed that there were no differences in the likelihood of attacking between in- and out-group opponents in equal and weak conditions. However, the rate of attack on out-group members exceeded the rate of attacks on in-group members only in the strong condition.

The results from the two studies were inconsistent. Meta-analytic results and future directions are discussed.

Short session 1 | Room 1

Tuesday July 2, 2024 09:10 - 09:15

**Fairness in the face of changing inequality**

Dianna Amasino | Rafaela Pinto

Tilburg University | Laterite

We explore how varying the level of inequality impacts self-serving allocation biases. One crucial factor in allocations is whether inequality is perceived as due to merit or luck, with inequality due to merit more accepted. The contributions of merit and luck are often hard to disentangle in society, but we use a controlled setting in which merit and luck are explicit and luck impacts the level of inequality. Past literature finds that those advantaged by luck allocate a higher proportion of joint earnings to themselves. We build on this work by examining the impact of varying inequality due to luck; higher inequality could be used by the advantaged to justify keeping even more for themselves or it could backfire by increasing the prominence of luck and undermining self-serving justifications, reducing allocation biases. We preregistered our hypotheses: https://aspredicted.org/67M_TVY. All participants (N=400) completed real-effort tasks, after which they were split into recipients and dictators. Dictators were matched with different recipients over 30 rounds and chose how to divide the pair's joint earnings. Joint earnings were calculated by multiplying the number of correct answers (merit) by a pay rate (luck) that was randomly assigned to be high or a low, with pairs of dictators and recipients always having opposite pay rates to create inequality. Finally, the ratio between high and low pay rates varied in blocks within-participants, with the advantaged earning 3, 7, or 15 times the disadvantaged per correct answer, presented in either increasing or decreasing order. In multilevel regressions of the percentage of joint earnings kept, we find that those advantaged by a high pay rate keep more for themselves ($b=10.3$, $p=0.001$), controlling for task performance and individual factors, suggesting that inequality due to luck is used as a justification to keep more. Further, we find that the advantaged keep significantly more at higher pay ratios ($b=0.45$, $p<0.001$), although not in proportion with the increasing pay ratio, providing more support for the view that luck is used by the advantaged to keep more for themselves but not to the full extent. Further, in exploratory analyses, we find that the direction of inequality matters, with advantaged responding more to pay ratio when inequality starts high and decreases than when it increases from a lower level ($b=-0.34$, $p=0.006$), suggesting that initial inequality and the direction of change are important.

Short session 1 | Room 1

Tuesday July 2, 2024 09:15 - 09:20

**Relationship Interdependence Influences Prosocial Behavior Across Ego Networks**

Daniel Balliet | TIffany Matej Hrkalovic | Francesca Righetti

Vrije Universiteit Amsterdam | Vrije Universiteit Amsterdam | Vrije Universiteit Amsterdam

People are embedded in dense social networks characterized by a diverse set of relationships, which can vary tremendously in interdependence – that is, how each person's actions affect each other's outcomes. To date, there is no unifying framework of how people evaluate their relationship interdependence. We propose that people evaluate their relationships along several dimensions of interdependence including mutual dependence, power, conflicting interests, information certainty, and anticipated interdependence.

We had participants (N = 207) from an online panel (Flycatcher) use these items to describe diverse relationships in their network. Participants reported on 10 relationships within their network that were high/low on each dimension of interdependence (5 dimensions x 2 level (high, low)). They completed a set of questions about the reported person (i.e., age, sex, closeness), and then completed the Relationship Interdependence Scale. Finally, they participated in two modified dictator games: (1) N-person Dictator Game (DG) in which they distributed 100 lottery tickets among themselves and 10 people they previously reported on and (2) a standard DG in which they were given 100 lottery tickets for each person and distributed the tickets between themselves and the other person. Each ticket had a chance to win a 50 euro gift card for one of two online shops in the Netherlands (Bol.com and CoolBlue).

We found that people evaluated the relationships along the five dimensions of interdependence, and that people could use the Relationship Interdependence Scale to differentiate relationships that differed on each dimension of interdependence. We found support for several pre-registered hypotheses that relationship interdependence could predict the relationships people allocated more tickets to in the Dictator Games. Relationship interdependence could explain up to 33% of variance in behavior in the Dictator Games. As expected, people allocated more tickets to individuals with whom they shared relationships higher on mutual dependence, information certainty, anticipated interdependence, and lower conflict of interests. Contrary to our expectations, relative power was not a significant predictor of allocation behavior. In conclusion, relationship Interdependence provides a multi-dimensional framework for understanding how people evaluate the diverse relationships in their social network, with consequences for understanding which relationships people invest in with prosocial behaviors.

Short session 1 | Room 1

Tuesday July 2, 2024 09:30 - 09:35

**Reputation-Based Trust and Cooperation: When and How Does Dishonest Reputation Upgrading Backfire?**

Yanyan Chen | Junhui Wu | Yugang Li | Baizhou Wu | Shenghua Luan

CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences | CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences | CAS Key Laboratory of Behavioral Science, Institute of Psycho

Reputation is a powerful mechanism for promoting trust and cooperation. Yet, its efficiency hinges on the prerequisite that reputation can accurately represent an individual's behavioral history. In real-life situations, people often use dishonest tactics, such as online shops providing cash incentives for positive ratings, to upgrade their reputation and reap reputational benefits. Despite the tremendous efforts dedicated to detecting and reducing dishonest reputation upgrading behavior, it remains unclear whether reputation-based cooperation would collapse when dishonest reputation upgrading occurs in a reputation system and how this behavior affects trust and cooperation. To address these questions, we conducted three studies. In each study, participants interacted with high- and low-reputation targets in reputation systems with or without opportunities for dishonest reputation upgrading, and participants' trust and cooperation toward the targets were measured. In the first two studies, hypothetical scenarios involving interpersonal interactions (Study 1, N = 163) and hotel reservations (Study 2, N = 200) were used. In Study 3 (N = 252), we focused on actual cooperative behavior with real monetary incentives and sought to distinguish the potential roles of integrity-based trust and benevolence-based trust in explaining the effects of dishonest reputation upgrading on cooperation. Overall, the results consistently showed that whereas dishonest reputation upgrading undermined cooperation with high-reputation targets, it paradoxically promoted cooperation with low-reputation targets, and trust toward the targets played a critical mediating role in explaining these opposing effects. These findings suggest that opportunities for dishonest reputation upgrading do not bring about the anticipated indirect benefits for high-reputation partners but instead inadvertently benefit those with lower reputations. To tackle challenges posed by dishonest reputation upgrading, we call for governments and policy makers to engage in intervention efforts that make reputation sharing and transmission more transparent and strictly sanctioning those who use deceptive strategies to upgrade their reputations without being cooperative themselves.

Short session 1 | Room 1

Tuesday July 2, 2024 09:35 - 09:40

**Does Realism Affect Behaviors in Moral Dilemmas?**

Matthieu Légeret | Laetitia Renier | Emmanuelle Kleinlogel

University of Amsterdam | Université de Lausanne | Université de la Réunion

Moral dilemmas are often studied through hypothetical scenarios. For instance, the famous trolley dilemma (Thomson, 1985) consists of describing to participants a trolley heading towards five people walking on the tracks. The participants have to choose whether to pull a lever, redirecting the trolley towards one person walking on a different set of tracks. In this scenario, acting is considered to follow a utilitarian approach to morality, as it saves five lives by sacrificing one. However, hypothetical scenarios have been criticized in the past for their lack of realism (Hughes & Huby, 2002; Roehling, 1999). Virtual reality (VR) has been proposed as a more immersive and, therefore, realistic alternative to study moral reasoning (e.g., Parsons, 2015; Parsons et al., 2017).

To investigate how realism affects moral decisions in dilemmas, we conducted a preregistered laboratory experiment (n = 293) in which we implemented three versions of the trolley dilemma. Building on the research by Patil et al. (2014), we operationalized the trolley dilemma as a vignette (Text condition) as well as in a virtual reality (VR) environment. To further study how realism impacts decisions, we manipulated the level of graphics in the virtual reality scenario (VR Low versus VR High). We depart from Patil and colleagues in that we measured behaviours in all versions of our experiment, instead of measuring judgments in the Text condition. Furthermore, we conducted the experiment as a between-subject experiment, randomly allocating participants to either the Text, VR Low, or VR High conditions. Furthermore, we measured personality traits as well as moral disengagement and controlled for the participants' attitude towards technology.

Our results show that decisions in the text and VR (High and Low) scenarios do not differ significantly. Likewise, the increase in realism within the VR conditions did not significantly impact the decisions of the participants. However, we find that Honesty (from the HEXACO personality questionnaire) has significantly different effects on the decision to pull the lever in the Text and VR conditions. Whereas past research has found a negative relationship between honesty and utilitarianism in moral dilemmas (e.g., Djeriouat & Trémolière, 2014), we find that honesty does not predict decisions in the Text condition. However, we find that honesty positively predicts utilitarian decisions in the two VR versions of the dilemma.

Short session 1 | Room 1

Tuesday July 2, 2024 09:40 - 09:45

**Reputation dynamics in divided societies: image updating in indirect reciprocity under private assessment**

Isamu Okada

Soka Univeristy

Exploring the evolutionary mechanisms of cooperation in societies where reputational consensus cannot be expected, as assumed in divided societies, is crucial for understanding the fundamental principles of human behavior in modern societies. While indirect reciprocity serves as a major explanatory mechanism, existing studies predominantly concentrate on the assessment of donors' images. Limited attention has been given to scenarios where different individuals disagree on who deserves punishment.

In this talk, we present an agent-based model facilitating the updating of both donors' and recipients' images. Our comprehensive simulations reveal that commonly analyzed assessment rules, focused on updating donors' images, rank as the second-best option. In contrast, an assessment rule updating bad images is found to be the most effective in sustaining cooperative regimes. Specifically, when updating the image of either a donor or a recipient becomes necessary, adopting an assessment updating rule that alters the image of a person with a bad image is advisable, irrespective of their role as a donor or a recipient.

This study identifies a social norm prioritizing a good image, characterized as tolerant. Such a norm aligns with previous research emphasizing the significance of tolerant evaluation in private assessment schemes.

Short session 1 | Room 1

Tuesday July 2, 2024 09:45 - 09:50

**Pro-social and Self-serving default choices in the common pool resource dilemma are persuasive but not persistent**

Eladio Montero-Porras | Rémi Suchon | Tom Lenaerts | Elias Fernández Domingos

Vrije Universiteit Brussel | Université Catholique de Lille | Université Libre de Bruxelles | Université Libre de Bruxelles

Extracting from Common Pool Resources (CPR) requires making choices to balance personal profit and sustainability. Therefore, the way these decisions are structured is vital for developing policies that encourage sustainable resource use. We conducted an experiment to explore how default settings in decision-making can influence sustainable resource use. We tested whether defaults can nudge cooperative individuals to extract more and selfish individuals to extract less, contributing to sustainable resource governance. In our experiment, participants (n=412) played a CPR dilemma game in groups, with varying default extraction values. The game had three treatments: a pro-social default (n = 156), a self-serving default (n=156), and a control with no default (n=100). The defaults were removed after five rounds to study their long-term effects. Moreover, we measured participants' social preferences with the Social Value Orientation and Risk Assessment tasks.

We found that the self-serving default led to increased extraction for five rounds with respect to the average extraction in the control treatment, where no default was shown, while the pro-social default resulted in reduced extraction for three rounds, with respect to the control treatment. The default effect was asymmetrical - the impact of a selfish default lasted longer than a pro-social default (5 and 3 rounds). After we removed the default value, no significant differences were found among the treatments, indicating no long-term impact. Notably, defaults influenced participants with conflicting inclinations; cooperative individuals (as measured by the SVO) extracted more with the self-serving default and selfish individuals extracted less when facing a pro-social default.

In conclusion, the results presented in this word helped us understand some of the factors - besides economic incentive - play a role in how we make decisions. We showed how participants can essentially be nudged into a certain scenario by changing a value. This has implications on the way we consume finite resources and how we interact with technology. In the case of systems where a default value has to be enforced by design, or they can be an option, defaults make the decision-making process easier to pick the option in which is the best interest for societal good. On the other hand, bad defaults proved to be a mechanism that cooperative participants used to extract more on average. While good defaults can nudge otherwise selfish people, bad default have the potential to do the same with a greater extent.

Short session 2 | Room 1

Wednesday July 3, 2024 15:45 - 15:50

**No “cooperation for reputation” in highly asymmetric step-level public goods games**

Yukari Jessica Tham | Yohsuke Ohtsubo | Kaori Karasawa

Kobe University | The University of Tokyo | The University of Tokyo

People contribute more to public goods when their reputation is at stake (Wu et al., 2016), which always enhances collective interests in linear public goods games. In contrast, in step-level public goods games, such an increase in contribution may enhance or undermine collective interests, which depends on the balance between a reduction in “under-contribution” and an increase in “over-contribution.” A previous study (Hardy & Van Vugt, 2009) examined this possibility using vignettes depicting step-level public goods games. However, it has not been examined in games involving actual interactions among participants.

To address this gap, we conducted an online step-level public goods game experiment, comparing public and private conditions. In both conditions, participants (N = 332) were matched with another participant and played a volunteer’s dilemma game (the simplest step-level public goods game; Diekmann, 1985) for 24 rounds. In each round, participants independently and simultaneously decided whether to volunteer and incur some costs. If at least one participant in a dyad volunteered, two received 80 coins each. While dyads in the public condition saw each other at a post-game video meeting, those in the private condition did not have such a chance. We examined three types of payoff structures. In symmetric games, the volunteer cost was 50 coins for both players. In moderately asymmetric games, it was 50 coins for one player (“weak player”) and 30 coins for the other player (“strong player”). In highly asymmetric games, it was 50 coins for one player (“weak player”) and 10 coins for the other player (“strong player”).

Results of the symmetric and moderately asymmetric games were in line with the predictions; they showed significantly less under-contribution and slightly more over-contribution in the public condition than in the private condition. As a result, dyads earned higher payoffs in the public condition. Contrarily, there was no difference in these variables between the public and private conditions in the highly asymmetric games. Specifically, strong players volunteered in most rounds, while weak players rarely volunteered, no matter whether their reputation was at stake. These results indicate that people do not always increase their contribution when their reputation is at stake. Specifically, if the role of each group member is clearly defined (as in the highly asymmetric games), people seem unwilling to deviate from their roles.

Short session 2 | Room 1

Wednesday July 3, 2024 15:50 - 15:55

**Intergroup vicarious retribution caused by prosocial punishment in social dilemmas**

Ryoichi Onoda

Department of Sociology, Daito Bunka University, Japan

Rationale: Previous research on social dilemmas has consistently shown that punishing defectors increases cooperation. Consequently, punishment is considered a prosocial behavior undertaken for the collective benefit (Fehr & Gächter, 2002). In this study, I imagined a situation in which two groups existed and were free to punish all others. Accordingly, some findings on aggressive behavior allowed me to predict that punishment would produce new negative phenomena.

Nawata & Yamaguchi (2013) and Kumagai (2013) documented that when an out-group member (A) engaged in aggressive behavior toward an in-group member (B), an individual (C) in the victim's group tended to retaliate against the aggressor (A) or the aggressor's group member (D). This retaliatory behavior is called “intergroup vicarious retribution”. Huang et al. (2015) showed that perceived threats to the out-group might cause the attribution of aggressive intentions of the out-group, leading to aggression against the out-group. These findings predict that when punishment is imposed by an out-group member, individuals who feel threatened by the out-group perceive the aggressive intentions of the out-group and engage in intergroup vicarious retribution.

Methods and Results: Initially, I conducted a laboratory experiment to compare conditions in which an out-group member (A) punished an in-group defector (B) (49 participants) with conditions in which a randomly selected in-group defector (B) was punished by the computer (41 participants). The purpose was twofold: to find out whether intergroup vicarious retribution occurred (C to A or D) and reveal the psychological underpinnings of this behavior. As a result, intergroup vicarious retribution by participant (C) against the out-group punisher (A) occurred. Regarding psychological underpinnings, the "awareness that intergroup vicarious retribution is praised by in-group members" was a more important factor than the out-group threat.

In the second experiment, I compared conditions in which an out-group member engaged in aggressive behavior toward an innocent in-group member (42 participants), conditions in which an out-group member punished the in-group defector (41 participants), and conditions in which an in-group member punished the in-group defector (43 participants) to determine the subjective evaluation of punishment by an out-group member. Results indicated that punishment by an out-group member was rated as more prosocial and less malicious than aggressive behavior and more similar to punishment by an in-group member.

Conclusions: Intergroup vicarious retribution can generate a chain of retaliations involving innocent bystanders, developing into a heinous macro-phenomenon. This study provides new evidence that prosocial punishment in social dilemmas causes a more heinous macro-phenomenon than previously thought.

Short session 2 | Room 1

Wednesday July 3, 2024 15:55 - 16:00

**Wanna solve the climate crisis? End inequality. – The potential of unequal loss in a climate game kills trust and causes sinking together.**

Judit Mokos | Hubert Janos Kiss | Zsóka Vásárhelyi | Adrienn Král | Zoltán Kovács | István Scheuring

Eotvos Lorand University, Budapest | Institute of Economics, Centre for Economic and Regional Studies; Department of Economics of the Corvinus University, Budapest | Evolutionary Ecology Research Group, Institute of Ecology, Budapest | Cooperation and Evo

Rationale: The main difficulty with climate change is coordinating an international action, where countries are also competitors in the global economy. The temptation of not investing enough makes climate change negotiations a good example of a common pool resource dilemma. Less wealthy countries and individuals suffer the consequences of climate change more pronouncedly. Monetary inequality and the competition caused by it could undermine the possibility of cooperation.

Method: Using a modified climate game, we studied how pronounced competition among group members affects the willingness to contribute to the cooperative effort. To model the unequal consequences of unequal wealth in the treatment group if the threshold was not reached, participants lost different percentages of their funds: players with a higher amount of spare money lost less than players with the lowest sum in their pocket. In control groups, everybody's loss was the same in case of not reaching the threshold. If the threshold is reached, everyone receives 100% of their remaining endowment. 12-12 groups of six people were tested in both treatments. Two weeks before the game participants filled out a series of psychological questionnaires measuring their altruism, climate attitude, and risk-taking. For more details, see the preregistration of the study: https://osf.io/rhky4/

Results: The competition resulted in less cooperation straight at the beginning of the game, after that the players reached the threshold with a lower chance than groups with no competition. Participants contributing more seem to be more altruistic in other situations as well. Even though the game did not mention the word "climate", participants who contributed more tend to believe in climate change and act against it more.

Conclusion: Our result is consistent with previous studies showing that first impression defines future cooperation. It also highlights the importance of inequality in cooperation.

Overall, the potential of unequal loss kills the trust in your fellows and could cause them to sink together.

Short session 2 | Room 1

Wednesday July 3, 2024 16:00 - 16:05

**What kind of information are people willing to spread?: Manipulating information credibility to examine information transfer bias in situations of indirect reciprocity.**

Rie Mashima | Itsuki Kawamura | Nobuyuki Takahashi

Health Sciences University of Hokkaido | Hokkaido University | Hokkaido University

Theoretical studies have concluded that the key to the emergence of indirect reciprocity is discriminate altruism based on 1st-order information (others' previous behaviors) and 2nd-order information (reputation regarding targets of others' previous behaviors). Those studies assumed the possibility of perceptual errors but did not consider any bias that skews disseminated information in a specific direction. In reality, however, people may have a bias in telling others what they have observed. If people have such a bias, the circulated information in society would be skewed, which calls for reexamining the theoretical conclusion regarding the emergence of indirect reciprocity. Takahashi et al. (2022) showed empirical evidence of such a bias using a vignette. However, the robustness of their findings has not been established. Thus, the primary purpose of the current study is to reexamine information transfer patterns by conducting a conceptual replication of Takahashi et al. (2022). The secondary purpose is to explore the possibility that information credibility affects information transfer patterns.

We conducted a vignette study. Respondents (N=1218, undergraduate students) read a scenario in which they observed the behavior of one person (a donor) to another person (a potential recipient). After reading a scenario, respondents reported their intention to tell the donor's behavior to others. We manipulated 4 between-subjects factors: (a) the donor's behavior (cooperated or defected), (b) the reputation of the potential recipient ("good" or "bad"), (c) the donor's reputation before s/he behaved ("good" or "bad"), and (d) information credibility (high/low).

Results showed that respondents were less willing to spread the behavior of the donor whose reputation had been "good" than that of the donor whose reputation had been "bad." We also found that respondents were less willing to spread the donor's defective behavior than the donor's cooperative behavior. These results replicated those of Takahashi et al. (2022). Information credibility did not affect information transfer patterns.

Our results suggested that people are biased in spreading what they have observed to others. People seem to have a tendency to refrain from spreading out the information regarding 1) others who have a good reputation and 2) other's defective behaviors. These results suggest that the information available in our society is skewed because of peoples' bias in transferring information in situations of indirect reciprocity. Thus, it may be necessary to build a new model of indirect reciprocity that incorporates peoples' bias in spreading others' information.

Short session 2 | Room 1

Wednesday July 3, 2024 16:05 - 16:15

## Are Christians More Forgiving and Less Greedy? Evidence from a Power-to-take Game Experiment

Bing Jiang

Virginia Military Institute

A substantial amount of literature has demonstrated that religious beliefs and practices foster prosocial attitudes and behaviors such as generosity, altruism, cooperation and care for others. Despite promising advancements of knowledge in understanding the relationship between religion and prosocial attitudes and behaviors, little effort has been made to study how religion is linked to negative reciprocity and antisocial behaviors. In this paper, I investigate whether and how Christian belief is linked to decision-making in a two-player power-to-take game experiment. I recruit 714 participants from Amazon Mechanical Turk (Mturk) to conduct an online power-to-take game experiment combined with surveys. I find that overall, participants who possess a genuine Christian belief tend to take less resources and also destroy less resources when exposed to potential resource extraction from others than those who do not have the Christian belief, regardless of their counterparts' religious background - that is, Christians are indeed less "greedy" and more "forgiving" than non-Christians. Interestingly, the trait of negative reciprocity is statistically significant in explaining antisocial interactions: participants who score high on negative reciprocity are more likely to take resources from others and destroy their own resources. These findings have implications for understanding the effect of religious beliefs on decision-making and antisocial behaviors.

Short session 2 | Room 1

Wednesday July 3, 2024 16:15 - 16:20

**How do Impressions of Age, Ethnicity, Sex, and Social Class Simultaneously Affect Cooperation?**

Paul A. M. Van Lange | Josh Tybur | Lei Fan | Niels van Doesum

Vrije Universiteit Amsterdam | Vrije Universiteit Amsterdam | Vrije Universiteit Amsterdam | Leiden University

Rationale: When we meet new people, more than one impression may meet the eye. For example, imagine you are in Leiden just wondering around. Upon meeting another person, you see their sex, ethnicity (a person from Suriname, China, Turkey, or The Netherlands), have a clear impression about their age (say around 25 or around 60 years), and even about their social class (low versus middle/high). The most well-observable cues such as sex and ethnicity (and to some degree, age) have received considerable attention in the literature, especially in the literature on impression formation. But impressions of another's social class have received hardly any attention. And we do not know of any research that examined these attributes simultaneously, which is actually happens most often in everyday life in Leiden and elsewhere. Thus, the goal of the present research is to examine the simultaneous influence of these impressions on human cooperation, thereby using various measures to provide a bigger picture and address generalizability across low-cost (social mindfulness) and high-cost (social value orientation) cooperation.

Methods: A total of 7,069 participants (MAge = 55.11, SDAge = 16.6, NFemale = 4,447) saw a picture of an individual and read a description of that person. Pictures were obtained from existing databases that portrayed a person who was either male or female, young or old, and White, Black, Asian, or Turkish. The old faces were manufactured using FaceApp, and the young faces were original faces from the face databases. Descriptions included the target's name (prototypically Dutch for White faces, Surinamese for Black faces, Chinese for Asian faces, and Turkish for Turkish faces). Social class was manipulated within targets, with verbal information.

The dependent measures were prosociality assessed with the Slider Method and social mindfulness, to capture high-cost and low-cost cooperation.

Results: Findings provided support for a fairness perspective, such as that a person from lower social classes elicited greater cooperation than those from higher social classes. An unexpected finding was that ethnic minorities also elicited greater cooperation. No strong effects for age or gender were found. These findings were supported in both within- and between-participant analyses.

Conclusions: People may view members of groups that seem disadvantaged as more strongly in need of cooperation. Fairness considerations may underlie such behaviors, perhaps coupled with the notion that they feel that White people from higher social classes are already privileged and therefore less deserving.

Short session 2 | Room 1

Wednesday July 3, 2024 16:20 - 16:25

**Managing the Commons: The Role of Political Orientation and Framing on Cooperation in a Common-Pool Resource Dilemma**

Richardt R. S. F. Hansen | Bryony Buck | Johannes A. Koomen | John Betts | Ana-Maria Bliuc | Rebecca M. Koomen

University of Dundee | University of Dundee | University of Rochester (Ret.) | University of Monash | University of Dundee | University of Dundee

Rationale: Cooperation between political parties is essential for negotiating policies that determines the trajectory of global social dilemmas such as climate change. Previous research suggests differing political orientations affect social decision-making insofar that liberals cooperate more than conservatives. Further, cooperation in social dilemmas is influenced by the framing of the social dilemma. Whether political orientations moderate framing effects in resource dilemmas remains an open question. This research investigated whether Democrats and Republicans experienced framing effects equally in a resource dilemma. It was hypothesized that (i) Democrats cooperate more than Republicans, (ii) Framing the resource dilemma environmentally results in more cooperation than framing the dilemma neutrally, and (iii) political orientations moderate the existence and magnitude of framing effects. Understanding this moderation effect will assist policymakers and researchers to inform policymaking and guide interventional strategies.

Methods: Prolific.org was used to recruit 266 American participants identifying as Democrats and Republicans. A 2x2 between-subjects design was utilized with political orientation (Democrat, Republican) and framing (environment, neutral). The pre-registered experimental procedure followed three steps: (i) Democrats and Republicans randomly allocated to the environmental or neutral framing condition, (ii) Complete political orientation survey, (iii) Play the online resource dilemma game, framed environmentally (“environment game”) or neutrally (“fishing game”).

Results: Framing the resource dilemma environmentally resulted in more cooperation, both initially and overall. Further, political orientations predicted cooperation: Democrats initially cooperated more than Republicans, but not across the entire resource dilemma. Political orientation was not found to moderate the existence nor magnitude of framing effects in the resource dilemma. Yet, further exploratory analyses suggest that strength of political orientations should be considered. Here, political orientations do moderate the existence and magnitude of framing effects insofar that framing effects exist only for strong Democrats.

Conclusions: While findings did not support the hypothesis that political orientation moderates framing effects in a resource dilemma, framing effects do exist insofar that framing the resource dilemma environmentally results in more cooperation than framing the dilemma neutrally. Consequently, framing small-scale – and possibly large-scale – resource dilemmas environmentally may serve as a useful strategy to increase cooperative behaviours. Likewise, political orientation did influence cooperation initially in the resource dilemma insofar that Democrats cooperated more than Republicans, suggesting that Democrats are more likely than

Republicans to engage initially with cooperative behaviours. Whilst the exploratory analyses suggest that strength of political orientation moderates framing effects, it is up to future research to confirm.

Short session 2 | Room 1

Wednesday July 3, 2024 16:25 - 16:30

**Social Corrections: Nudging Norm Enforcement Against Fake News Sharing**

Eugenia Polizzi | Giulia Andrighetto | Amalia Alvarez-Benjumea | Biljana Meiske

Institute of Cognitive Sciences and Technologies, National Research Council of Italy | Institute of Cognitive Sciences and Technologies, National Research Council of Italy | Institute of Public Policies and Goods, National Research Council of Spain | Euro

Corrective comments, left by users as responses to posts containing inaccurate information, not only effectively reduce belief in misinformation among those observing the interaction, but also serve as a publicly observable punishment against violations of norms regulating online content sharing. By visibly showcasing punishment, corrective comments hold the potential to operate as a norm-nudge, updating observers' perception of the norms regulating punishment, so-called "meta-norms". Specifically, we posit that witnessing social corrections provides information about how frequent and socially appropriate other users find punishing norm violations, and that stronger meta-norms increase the likelihood that potential enforcers will engage in corrections when needed.

To empirically test this hypothesis, we conducted a pre-registered online experiment inviting 660 participants to engage in discussions resembling an online forum. Participants were free to comment on posts shared by prior users. We experimentally manipulated whether participants could observe corrections left by others in response to posts featuring inaccurate information. We then examined the effect of this variation on both behavior – the likelihood of replying with a corrective comment- and norms -the perceived social appropriateness of correcting inaccurate posts. Norms were elicited using standard incentivized procedures.

We show that participants exposed to posts corrected by other users were significantly more likely to reply with a corrective comment, even after accounting for participants' perceived accuracy of the shared post. Importantly, participants exposed to corrections also perceived replying with corrections to be more socially appropriate compared to control subjects, supporting the idea that social corrections can act as a norm-nudge at the level of punishment ("meta-norm" nudge).

These findings complement research on the role of social norms in influencing people's willingness to engage in punishment. Indeed, we show that exposure to punishment can lead to a shift in meta-norm perception and foster norm enforcement. From a policy perspective, our results suggest that social corrections are a promising approach to encourage users to speak out against fake news in online contexts, and that interventions targeting potential enforcers can be a potent tool to complement policies specifically directed at norm violators.

Poster session | Poster room

Tuesday July 2, 2024 16:30 - 16:35

**Learning the value of Eco-Labels: The role of information in sustainable decisions**

Alejandro Hirmas | Jan Engelmann

Universiteit van Amsterdam | Universiteit van Amsterdam

The European Union aims to become carbon-neutral by 2050. One of their new policies is to develop a sustainability rating to promote more sustainable consumption. However, how these ratings will affect consumers' decisions is unknown. We design an incentivized experiment in which participants make multiple decisions between two artificial products that vary in price, quality, and sustainability. The quality and sustainability levels of the products are presented in ratings, and we test the impact of different underlying rating systems in different treatment groups. In the middle of the experiment, some groups learn that both ratings increase linearly, while other groups that a linear increase underlies one rating while a convex increase underlies the other rating. This information treatment allows us to explore how different types of ratings affect the decision. In addition to the choice data, we also record the information search (i.e. visual attention) to better understand individual differences in the participants' decision processes. We find that if the consumers have no further information about the rating, consumers treat quality and sustainability ratings similarly, but when we incorporate the attention data, we find distinctive patterns of how quality and sustainability are approached. Moreover, after participants learn more about the ratings, their behaviour further differs between both ratings. When some sustainability products are less valuable (than what they expected), participants shift to both lower- and higher-sustainability products. In contrast, when the quality of a product becomes unappealing, we only observe shifts towards high-quality products. We conclude that ratings informing regarding their own benefits, i.e. quality, might be treated differently from ratings regarding sustainability, and thus need to be carefully explained/designed to promote more sustainable consumption.

Poster session | Poster room

Tuesday July 2, 2024 16:35 - 16:45

**Is cooperation sustained under increased mixing in evolutionary public goods games on networks?**

Wei Zhang

ETH Zürich

Well-mixed populations model population-wide interactions since everyone shares the same set of interacting partners, the entire population. While networks model local rather than global interactions by restricting them to social neighborhoods. Therefore, a question arose: when individuals interact in groups on networked populations, if there is a probability to connect two groups and form an additional global group, whether the additional mixing links always play a positive or negative role in the evolution of cooperation?

In this work, we propose an evolutionary game model that is able to capture the effect of long-range links mixing local neighborhood and global group interactions in a finite networked population. We derived dynamical equations for the evolution of cooperation under weak selection by employing the mean-field and pair approximation approach. Using properties of Markov processes, we can approach a theoretical analysis of the effect of the density of mixing links. We find a rule governing the emergence and stabilization of cooperation, which shows that the positive or negative effect of mixing-link density for fixed group size depends on the global benefit in the public goods game. With mutations, we study the average abundance of cooperators and find that increasing mixing links promotes cooperation in strong dilemmas and hinders cooperation in weak dilemmas. These results are independent of whether strategy transfer is allowed via mixing links or not.

The work is published in Applied Mathematics and Computation, 2023. https://doi.org/10.1016/j.amc.2022.127604

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Cooperative dilemmas in rational debate**

Shang Long Yeo | Julián Garcia | Toby Handfield

National University of Singapore | Monash University | Monash University

Rationale: Individuals engaged in a rational debate face a mixture of incentives for both cooperation and competition. We all benefit from collective achievements such as rapid consensus on true beliefs. But as individuals we also prefer not to change our minds. This study aims to use formal modeling to investigate whether these factors give rise to cooperative dilemmas, similar to those studied in other domains. We also use the framework to represent the politicization of a debate and use it to make predictions regarding the epistemic quality of debate under those circumstances.

Methods: We use an agent-based model, derived from the work of Gregor Betz. Two agents occupy positions in a space of logical possibilities. Arguments are introduced according to minimal rules of logical consistency, and debaters must revise their beliefs if their previously occupied position is removed by a valid argument. Debaters are motivated to find the truth but also to minimize the number of beliefs on which they change their mind. Roughly speaking, agents in the model can use "aggressive" strategies, which undermine their opponent's position; or they can use "defensive" strategies, which support their own position.

Results & Conclusions: The model predicts the existence of social dilemmas, in which the individually optimal argument strategy will often block the collectively optimal outcome. We then extend the model to describe what happens when debate becomes politicized. We show that politicization of a debate makes the social dilemma even more acute. We discuss possible interventions to improve the prospects for cooperative and productive debate.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Retrospective Self-Reported Adversities and Family Unpredictability Over the Course of Childhood Uniquely Predicts Cooperative Behavior during Adulthood: The Moderating Role of Sensitivity to The Environment**

Libera Ylenia Mastromatteo | Sara Scrimin

Department of of Developmental Psychology and Socialization - University of Padova | Department of of Developmental Psychology and Socialization - University of Padova

Cooperation is an adaptive prosocial behavior involving two or more individuals working together to achieve a shared goal. While extensive research has explored factors influencing cooperation, scant attention has been given to the impact of childhood environments. Childhood experiences significantly shape adult behavior and personality. Specifically, family unpredictability and exposure to adversities during childhood are two crucial factors influencing cognitive and behavioral development across the lifespan. Family unpredictability, characterized by inconsistent family-behavior patterns, has been associated with detrimental outcomes, such as risk-taking behaviors and present-focused decision-making. Adverse childhood experiences have been correlated with detrimental outcomes and behaviors.

The impact of adverse and unpredictable environments on cooperative behaviors remains debated, with some studies indicating decreased cooperation while others showing no effect or even increased cooperation. Individual characteristics such as how environmental stimuli are perceived and processed, as reflected in the Environmental Sensitivity (ES), might explain inconsistency in the association between childhood environments and cooperation.

In the present study, 1730 participants (55.26 % male, Mage = 31.48, SD = 13.27) filled in an online survey in which they were asked to take part in a Public Goods Game. After that, participants retrospectively self-reported on the childhood adversities and experienced family unpredictability. Two lines of hypotheses were tested: (1) whether and how exposure to adversities and family unpredictability were associated with cooperation and (2) whether ES moderated the relationship between these two constructs and cooperation.

Results showed that family unpredictability and exposure to adversities were uniquely associated with cooperation, even after controlling for demographic information and personal values. Specifically, greater exposure to adversities over the course of childhood increased cooperation, while family unpredictability was negatively related to cooperative behaviors, thus showing an inverse relationship. In addition, ES moderated the relationship between exposure to adversities and cooperation, such that highly sensitive individuals who grew up in more harsh environments display higher levels of cooperation. The moderating role of ES was not statistically significant in the link between family unpredictability and cooperation.

Taken together, these findings indicate that childhood experiences have a significant impact on cooperative behavior in adulthood. Moreover, how individuals perceive, and process environmental stimuli plays a crucial role in moderating this relationship.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**An agent-based model of stochastic punishment by authorized third-parties in public goods game: The interplay between different justice concerns and reputation-based migration**

Ge-yang Chen

Department of SociologyJeonbuk National University

The current paper addresses three issues in existing models of third-party punishment. First, punishment is idealistically deterministic and immediately implemented. These underlying assumptions create an overly efficient and stable system, not in accordance with real-world complexities (e.g., free-riders are not equally likely to be getting punished). Next, while punishment from the perspective of restorative justice has received minimal attention particularly in the field of agent-based modeling, the overall evidence from experimental and non-experimental studies suggests that third-parties are more compensation-oriented than is previously reported. Lastly, not a few models examine the role of reputation in social dilemma games with the option of punishment, but reputation-based probabilistic punishment is rarely brought up. On the other hand, third-party punishment on the basis of reputation is shown to be more powerful as the strength of social ties is higher and residential mobility is lower, but its effectiveness remains to be seen in highly mobile societies.

Taken together, we propose an agent-based model of non-deterministic, non-immediate delegated punishment in unstructured populations (e.g., the era of Great Migration). Players join in a one-shot public goods game and then designated leaders (e.g., Leviathans) use the reputation scores of targeted individuals not only to guide their movement but also to determine the intensity of punishment. Our punishment mechanism rests upon retributive justice (stochastic punishment only) or restorative justice (stochastic victim compensation). We consider reputation-based movement as informal sanctions. Game participants are assumed to update their strategies through payoff-based imitation. The experimental conditions vary according to the types of punishment and the presence or absence of “voting with one’s feet.”

First, the average contribution to the public account is significantly lower under a combination of stochastic punishment and reputation-based migration than that in the baseline condition of deterministic punishment. Second, even when agents randomly move without relying on reputation, the level of cooperation remains considerably higher through noisy punishment by authorities than that in the absence of punishment. Nonetheless, such stochastic punishment only mitigates the declining rates of contribution. Third, unlike the scenario of random migration, when agents are allowed to move in specific ways guided by reputation, restorative punishment turns out to be more effective than retributive punishment especially in the latter half of simulation steps. Its compensatory aspect seems to enable collaborators to retain an advantage in preserving their strategies. Consequently, the number of cooperators is dynamically stabilized, which fosters emerging societies conducive to sustainable contribution.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Machine-learning Approaches for Meta-analytic Estimates of Important Predictors: Analysis of Cooperation in Social Dilemmas**

Yasuyuki Kudo | Takeshi Kato | Giuliana Spadaro | Daniel Balliet

Hitachi, Ltd. | Hitachi, Ltd. | Vrije Universiteit Amsterdam | Vrije Universiteit Amsterdam

There is a large amount of literature of empirical research on cooperation in social dilemmas. However, most meta-analyses on cooperation focus on a specific sub-topic within the literature, such as sanctions, communication, personality, and gender. One limitation of this approach is that it cannot assess the relative importance of all possible cooperation predictors. If this knowledge could be acquired, many benefits would accrue, including the planning of interventions to promote cooperation more effectively and efficiently, and improved study designs to test theories of cooperation. In this study, we demonstrated how machine-learning approaches can be used to analyze the importance of predictors within an entire field.

We used the Cooperation Databank, a machine-readable dataset of 2,636 experimental studies on cooperation (1958-2017), to conduct a meta-regression to assess the relative importance of 104 cooperation predictors, including parameters of the experimental paradigm (e.g., group size, payoff structure, and repeated interaction) and sample characteristics (e.g., gender, age, and ethnicity). Rarely reported predictors and those with dependent relationships have the potential to introduce bias in the importance estimates. To obtain robust and accurate estimates for these predictors, we applied several machine-learning approached, including methods such as the grouped-permutation feature-importance and ensemble of nonlinear meta-regression models.

The analysis revealed that the top five most important predictors were preferences for conditional cooperation, motivational orientation, conflict index in the Prisoner's Dilemma (PD), marginal per capita return in the Public Goods Dilemma (PGD), and punishment. We also found that the important rankings of the predictors varied by dilemma paradigm, i.e., PD, PGD, and Resource Dilemma (RD), with PD and PGD having a high rank correlation (r=.85), while RD having moderate correlations with the other dilemma paradigms (r=.64). The ensemble of meta-regression models that enabled the above analysis was able to explain 85% of the heterogeneity variance in cooperation across treatments through the combination of all predictors.

Our analysis identifies several important predictors besides changes to the incentive structure, including how much people value others' outcomes and whether they expect others to cooperate. We also found that the predictors of behavior may differ across dilemma paradigms. This suggests that this could be due to fundamental differences in how the decisions are made (giving versus taking) and the incentive structure. The machine-learning approach thus overcomes the methodological limitation that conventional meta-analyses focus on addressing a very specific phenomenon within a small portion of the literature.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**A framework to explore social norms in co-management of natural resources in a multi-level structure**

Caetano Franco | Eranga Galappaththi | Eduardo S. Brondizio | Michael G. Sorice

Virginia Tech - USA | Virginia Tech - USA | Indiana University - USA | Virginia Tech - USA

Social norms play a fundamental and often unobserved role in maintaining cooperative relationships and coordinating collective action, because they guide decision-making. We propose a cross-cultural framework that integrates the perspective of social norms in a multi-level structure. The conceptual framework begins by defining social norms and distinguishing what they are not. It explores the multi-level operation of social norms and emphasizes a cross-cultural perspective, examining individual and collective levels to understand the role of social norms in co-management of natural resources. Finally, we propose a path forward for investigating norms. Our definition of social norms centers on social expectations. Individuals follow a behavioral rule if they believe many people in their reference network adhere to it and believe they should adhere to it. It is distinguished from individual choices independent of others' actions, such as collective custom, shared moral rules, and legal injunction. Contextual understanding is key, as behaviors deemed acceptable in one context may differ in another. Social norms operate on multiple levels and vary across situational contexts. Cultivating social norms requires region-specific, grassroots strategies, emphasizing the impact of peer behavior and proximity on norm compliance. Promoting social norms requires acknowledging their interdependence from individual to group levels, as approval and sanction extend from individuals to communities. At the scale of the group or community, we consider the strength of the norm and tolerance for deviation to understand the role norms play in group dynamics. Tight cultures have strong norms, expect high conformity, and have little tolerance for violations of expected behavior. In contrast, loose cultures have more flexible expectations and greater tolerance for deviance. At the individual level, the forces that influence adherence to norms are articulated. This emphasizes the role of descriptive norms, what people do, and injunctive norms, what people approve of or disapprove of in influencing individual actions. It also looks at the dynamics of norm change and how individuals respond to shifts in these norms. For future research on social norms in co-managing natural resources, we propose: i) understanding norm dynamics within a cultural context; ii) investigating institutional interplay, governance, and conflict resolution; iii) examining communication and information flow in shaping norms; iv) emphasizing adaptive learning for continuous improvement in cultural and normative landscapes. Our cross-cultural framework offers insights into social norms, providing a strategic path for efficient co-management of natural resources in addressing and resolving social dilemmas.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**The moral dilemma and social dilemma of autonomous vehicles: The role of perceived responsibility**

Gary Ting Tat Ng

National Chengchi University

Rationale: When facing moral dilemmas, how should the algorithms of autonomous vehicles (AVs) be programmed? Past studies have documented that people in general approved of AVs that minimize injuries and casualties (utilitarian AV), but they preferred to purchase AVs that protect the passengers at all costs. This creates a social dilemma such that if the majority of AVs are programmed to protect the passengers at the highest priority, there would be more casualties. Past studies may have used ambiguous moral dilemma scenarios that confounded egoistic choice with deontological choice. That is, people chose to purchase AVs that save passengers may be adhering to the principle of deontology rather than being egoistic. In the present study, I created scenarios where both utilitarianism and deontology prescribed sacrificing passengers and examined how people make decisions in these scenarios. This study also tested how perceived responsibility of passengers and pedestrians can affect moral judgment and purchase intention of AVs.

Methods: A total of 402 participants were recruited through Prolific. Participants made decisions across 16 moral dilemma scenarios about an unavoidable accident occurring due to brake failure of AV. They were randomly assigned to either make moral judgment on how the AV should be programmed (moral judgment condition) or indicate their purchase intention of AVs programmed in different ways (purchase intention condition). For half of the scenarios, it was made clear that the pedestrians were innocent, while for the other half the situation was ambiguous. They also indicated the perceived responsibility of passengers and pedestrians for the accident in each scenario.

Results: Participants in the purchase intention condition indicated a higher tendency to sacrifice pedestrians than those in the moral judgment condition, replicating the past findings about the social dilemma of AV. More importantly, making clear that pedestrians were innocent reduced sacrificing pedestrians in the moral judgment condition, but not in the purchase intention condition. Furthermore, perceived responsibility of pedestrians predicted higher tendency to sacrifice pedestrians in the moral judgment condition, but had no significant effect in the purchase intention condition.

Conclusions: Taken together, these results suggested that when people consider which AV to purchase, they tend to ignore the perceived responsibility of passengers and pedestrians, and choose the one that protects passengers anyway. This study provides insights into how to solve the social dilemma of AV: making the responsibility of different road users salient may be effective in reducing the purchase intention of “egoistic” AVs.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Reciprocal Cheating in Children – Children Break the Rules to Return a Favor**

Laura Tietz | Felix Warneken | Sebastian Grueneisen

Leipzig University | University of Michigan | Leipzig University

Reciprocity, the exchange of favors over time, is a central mechanism facilitating human cooperation. From a young age, children reciprocate favors and other prosocial acts and even view doing so as obligatory (e.g., Warneken & Tomasello, 2013; Wörle & Paulus, 2019). In adults, however, reciprocal motives can also encourage ethically questionable behavior (e.g., when politicians return favors to campaign donors at the expense of their constituents and the general public; see e.g., Abbink et al., 2002; Lambsdorff & Frank, 2010). In two preregistered studies we investigated if the desire to reciprocate favors can encourage children, similar to adults, to break rules they otherwise approve of. In Study 1, 5-to-8-year-old children (n = 42) evaluated reciprocal and prosocial rule breaking in third-party contexts (i.e., story vignettes). Vignette protagonists either adhered to or broke the rules in order to repay a favor (reciprocity condition) or to simply benefit someone else (control condition; between-subjects design). We found that children condemned rule-breaking both when it was prosocially motivated and when it was a means of returning a favor. Rule-following was evaluated positively in both contexts. Thus, when assessing behavior of others, children prioritize rule adherence over returning favors. Study 2 investigated if, when faced with the decision themselves, 7-to-8-year-old children (n = 92) choose to adhere to or break the rules in order to return a favor. In a between-subjects design, children first interacted with a social partner wo either donated a valued resource to the child (reciprocity condition) or behaved neutrally (control condition). Subsequently, children had the opportunity to cheat in a game to benefit the partner. Using generalized linear models, we found that children transgressed significantly more in the reciprocity condition than in the control condition, and this tendency grew with increasing material indebtedness. Thus, in this first-party context children were willing to violate rules they otherwise support in favor of reciprocating prosocial acts (data collection with 5-to-6-year-old children is currently ongoing). These findings suggest that reciprocal motivations that are an important aspect of successful human cooperation can also compromise norm compliance and this can be observed already in childhood.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**The Effects of Emotions on Moral Reactions: A Meta-Analysis**

Xueting Zhang | Jaime Vigil-Escalera Sánchez | Niels J. Van Doesum | Lotte F. Van Dillen | Eric Van Dijk

Leiden University | Leiden University | Leiden University | Leiden University | Leiden University

Emotions have been demonstrated to shape people's moral reactions such as judging violations or punishing the violator. Despite a few reviews on the effect of disgust on morality (Donner et al., 2023; Landy & Goodwin, 2015), it remains unknown to what extent emotions as a whole shape moral reactions. In addition, the interchangeable usage of terminologies could be misleading. For instance, some studies asked participants "how morally wrong the violation is" to measure moral judgments, whereas other studies asked participants to choose if it is "permissible to kill one person to save five". Following moral philosophy, we refer to the former as descriptive judgments and to the latter as normative judgments. Moreover, research so far seems to over-focus on the roles of negative emotions in moral reactions (e.g., disgust, anger) while ignoring the more positive counterparts, especially the moral emotions (e.g., gratitude, awe).

To address the potential confusion and asymmetrical research interests, in this preregistered meta-analysis we aim to (1) summarize the strength of the relationship between emotions (as a whole) and moral reactions; (2) distinguish the influence of emotions on various types of moral reactions; (3) test several emotion- and morality-related moderators in the emotion-reaction link.

For the initial literature search, we utilized multiple techniques (e.g., wildcards and Boolean operators) to search PsycINFO, PsycArticles, Web of Science, and ProQuest. After deduplication, 3355 articles entered the abstract-screening phase, during which we excluded articles that did not meet our criteria and included 310 articles for the full-text screening. We expect to finish the screening before March 2024 and estimate to end up with 50—80 articles.

Considering the essential differences between moral responses, we plan to run three separate meta-analyses for descriptive judgments, normative judgments, and punishment, respectively. Because the same discrete emotions tend to have similar effects on moral reactions, we will fit three multilevel models where discrete emotions are at Level 3, studies are at Level 2, and individual effect sizes are at Level 1. We will perform subgroup analyses to address proposed moderators. For assessing study bias, we will employ a checklist for correlational studies (Cicolini et al., 2014) and the Cochrane risk-of-bias tool (RoB 2; Sterne et al., 2019) for randomized trial studies. For detecting publication bias, we will test the moderation effect of publication states and use techniques such as funnel plots and Egger's regression test (Egger et al., 1997).

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**When Ignorance is a Curse: Being Blind to Irrelevant Information Compromises Selection Decisions**

Hagai Rabinovitch | Yoella Bereby-Meyer

University of Amsterdam | Ben-Gurion University of the Negev

In France, it is legally prohibited to collect citizens' racial information without explicit consent, aligning with the government's commitment to "Color-Blind" principles. Similarly, the US Supreme Court has recently banned the use of race in university admissions. In some cases, being blind to information such as race and gender can reduce bias and improve fairness, but is it always the case?

Suppose interviewers tend to favor men over women during interviews, though gender is irrelevant to job performance. To account for the unjustified influence of gender, decision-makers (DMs) evaluating candidates based on their interview scores should adjust the score of male candidates by lowering it. However, using gender as a factor in the selection decision could be considered inappropriate and unfair, putting DMs in an uncomfortable situation. We suggest DMs would prefer to blind themselves to the biasing attribute, paradoxically believing they advance fairness while inadvertently compromising it.

Method. In three experiments (N = 559), participants read a scenario describing a selection for a position using a test that was affected by an irrelevant attribute (e.g., gender). We told participants that lowering the interview score of candidates who have unjustifiably benefited (e.g., men) leads to better decisions. We created a second scenario with a contextless irrelevant attribute ("Attribute X"), in which fairness considerations should be reduced. In Experiment 1, Participants rated how fair they believed the correction was. In experiments 2-3, participants had to choose between two candidates based on their test scores and were asked whether they wanted to know the candidates' values on the irrelevant attribute or to be blind to it.

Results. As hypothesized, participants rated the fairness of the correction lower for a personal irrelevant attribute (e.g., gender) than for a contextless one. Importantly, approximately 70% chose to be blind to the irrelevant attribute when it was personal, compared to only 45% when it was contextless.

Discussion. At times, DMs choose between candidates based on biased selection tests that disfavor women, minorities, or other groups. In order to address this issue, the DMs needs to adjust the test scores of those who were unfairly advantaged. However, such a correction is perceived as unfair and might put DMs in an uncomfortable position. People tend to believe they improve fairness by being blind to such information, when, in fact, they maintain a biased system and decrease accuracy.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Does social value orientation moderate the effect of reaction time manipulation on cooperation?**

Shogo Mizutori | Anri Ishida | Nobuyuki Takahashi

Hokkaido university | Hokkaido university | Hokkaido university

Studies focusing on reaction time (RT) have found that it exhibits an inverted U-shape relation with cooperation, i.e., a pattern in which extremely uncooperative or cooperative decisions are made fast and moderate ones are made slowly. Moreover, it has been reported that the moderation effect of social value orientation (SVO) underlies that relation. Specifically, there is a positive correlation between RT and cooperativeness among prosocials, while that among proselfs is null or negative. An explanation for the pattern is that fast decisions reflect their default responses, and prosocials' default is cooperation while proselfs is defection. For those data were obtained from self-paced studies, however, a more parsimonious explanation that attributes a long RT to the conflict between altruistic and selfish motivations would suffice. That is, given that those with either extremely selfish or altruistic preferences tend to experience little conflict and thus spend shorter time on decision-making, the inverted U-shape follows. Therefore, data that supports the explanation by the postulated difference in the default responses between SVO types need to be obtained from studies that manipulate RT. Since such studies are scarce, we conducted an experiment aiming to find if prosocial and proselfs respond differently toward RT manipulations.

Participants (128 undergraduate students in total) played a one-shot 4-person public goods game under one of the following 4 conditions: control, time pressure (TP), incentivized time pressure (ITP), and motivated delay (MD). In TP and ITP, participants were instructed to decide their amount of contribution as fast as possible, being presented with a countdown timer of 10 seconds. In ITP, participants could receive additional rewards depending on their decision time, while in TP, there was no incentive. In MD, participants had a chance to receive additional rewards by writing the reason for their decision before submission. After the game, they answered a post-experimental questionnaire including the SVO slider measure.

The results suggested that prosocials were more cooperative under ITP than the other three conditions. For proselfs, no obvious difference was found. The distribution of actually measured RT indicated that ITP was an effective manipulation to induce fast responses while TP was not, and it might explain the difference between these two TP treatments.

Although the results are consistent with the default-based explanation, the estimates involve large uncertainties due to the current study's small sample size. Thus, further data collection is required.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**The Impact of Between-Group Interaction on Within-Group Cooperation: a Meta-analysis**

Tycho van Tartwijk | Dan Balliet

Leiden University | VU Amsterdam

Cooperation, competition, and comparison are aspects of group behavior that have been well-researched in isolation, but the intersection of these elements remains relatively unknown. While competition and comparison between groups are both known to increase cooperation within groups, the precise differences between these factors are not clear. Furthermore, moderating influences on the relation between intergroup interactions and within-group cooperation have been left largely unexplored. Therefore, in this pre-registered meta-analysis, we sought to investigate the impact of between-group interactions, including cooperation and competition, on within-group cooperation. We conducted a systematic review of the literature comparing economic games with an intergroup interaction to single-group economic games across the PsycINFO and Web of Science databases. This produced a compilation of 15 relevant articles which yielded 47 effect sizes. Our multilevel random effects model revealed a small-to-medium, positive effect of intergroup interaction on within-group cooperation ($g = 0.505$), supporting past empirical research and evolutionary theory. Confirmatory moderator analyses indicated that males are more sensitive to intergroup comparison and competition, and thereby invest more in within-group cooperation during these situations than females. Additionally, competition was revealed to have a stronger positive effect on within-group cooperation than comparison. Exploratory moderator analyses of pertinent study characteristics highlighted that the recruitment method, game type, decision type, whether the game was repeated one shot, matching, deception, and whether participants were real or not had unique impacts on the relation between within-group cooperation and between-group interactions. Our findings present compelling meta-analytic evidence indicating the overall positive effect between-group interactions on within-group cooperation, which thereby can act as a possible solution to the free-rider problem. Additionally, the systematic review provides an account of past research in the field, which can aid future researchers in developing research methods and identifying gaps in the literature. Ultimately, while our research supports existing theoretical perspectives, mainly from an evolutionary standpoint, it also identifies certain discrepancies that may be addressed in future research.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Understanding exploitative behavior: Microtheory and evidence**

Hannes Rusch | Thomas Meissner | Gewei Cao | Maximilian Schmitt

Max Planck Institute for the Study of Crime, Security and Law, Freiburg; Maastricht University, Maastrichty | Maastricht University, Maastricht | Max Planck Institute for the Study of Crime, Security and Law, Freiburg | Max Planck Institute for the Study

Rationale: Exploitation is a global and perpetual societal problem. The Global Slavery Index, for example, counted 28 million people in forced labor in 2021. Accordingly, the UN identifies ending modern day slavery as a key target within their Sustainable Development Goal 8. While previous research has gathered rich data sets on the occurrence of exploitative behavior and its facilitation via human trafficking, we are lacking a deeper understanding of the common logic of situations which allow for exploitative behavior.

Methods: To close this gap, the individual incentives and situational determinants involved in exploitative interactions need to be analyzed. Here, we propose a game-theoretic model of the fundamentals of exploitative behavior. The model can be broadly applied to all situations in which party A proposes an interaction to party B and influences B's decision by coercive means. We derive its equilibria through backward induction and study its comparative statics, i.e., changes of the predicted equilibria induced by changes in the exogenous parameters of the model.

Results & Conclusion: Methodologically, our work contributes to principal-agent theory by relaxing the assumption that principals have no control over agents' outside options. Moreover, our model advances the study of exploitative behavior by yielding clear-cut predictions about individual behavior in situations affording exploitation which can be tested empirically. As a first step, we corroborate our comparative statics results empirically by testing them against current data sets on modern slavery.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Revealing the Interaction Between Strategic Properties and Intervention Opportunities: A Novel Taxonomy of Two-by-two Games**

Shacked Avrashi | Lior Givon | Ilan Fischer

University of Haifa, Israel | University of Haifa, Israel | University of Haifa, Israel

In game theory, a two-by-two game is one of the most basic and useful representations of strategic interactions between two parties, and is frequently applied to model social dilemmas. Classifying games according to their strategic properties can provide meaningful insights into the motivations driving the parties, reveal the predicted trajectories of the interactions, and in some cases also point to potential interventions aiming to influence the outcomes. Nevertheless, classifying games according to their payoffs neglects the role of the participants, specifically the interactions between the payoffs and the players. Furthermore, when considering the classification of games as a tool to influence interactions (e.g., increasing cooperation), a simple classification does not provide any guidance on how to achieve this goal.

Here we present a new taxonomy that merges two perspectives: (i) a revised version of Rapoport and Guyer's taxonomy (1966) that classifies all two-by-two games, and focuses on those categories that are of primary interest to social dilemma researchers; and (ii) a novel taxonomy that relies on the theory of subjective expected relative similarity, which addresses, not only the games' payoffs, but also the players' strategic perceptions of their opponents.

The revised Rapoport and Guyer taxonomy comprises five categories (absolutely stable, stable/ strongly stable, non-stable, prisoner's dilemma like, and no natural outcome), which address the expected outcome of the game and its stability. The second taxonomy, named the similarity-based taxonomy, comprises three categories (similarity-sensitive, one-player similarity-sensitive, and non-similarity-sensitive), specifying whether players' perceptions of their opponents have the potential to influence strategic decision-making. Specifically, the similarity-based taxonomy focuses on perceptions of strategic similarity, i.e., the likelihood of two players making similar choices.

The novel merged taxonomy comprises 15 game types. It addresses major strategic properties and allows forecasting expected trajectories. Most notably, the taxonomy points to social interactions that can be altered by influencing players' perceptions of each other.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Navigating Publication Bias in Judgment and Decision Making (JDM): A Philosophical Analysis of Scientists' Strategies and Perceptions**

Nora Hangel

Leibniz University Hannover; Leibniz Center for Science and Society, Institute for Philosophy

Rationale: When generating reliable results and effectively communicating them, researchers want to deliver meaningful contributions, maintain accountability, and attain recognition, for their academic viability. However, contribution, accountability, and the need for recognition form a social dilemma in scientific practice, a tension particularly pronounced in collaborative endeavors. Publication bias can lead to selective and subjective representation of research findings to increase publication success and can lead scientists to omit null findings or negative results and/or to overlook the importance of replication studies. The talk will contribute to understanding publication bias by analyzing scientists' reflections about evidence-checking practices and where to disseminate knowledge with varying degrees of confidence.

Methods: The study participants for this qualitative expert-interview study are experimental scientists from social psychology, behavioral economics, and others in the field of JDM. I use cognitive ethnography to study social epistemological processes, which are key to claims of objectivity, reliability, and empirical success in collaborative research. The naturalistic approach of the empirically informed philosophy of science project JUKNOW (The role of scientific judgment in generating knowledge) follows the understanding of being continuous with science by using science as a resource and conducting empirical investigation for a better understanding of social epistemological implications of scientists' practices.

Results: The first findings indicate a shift in how the field of JDM deals with forms of (scientific) uncertainties and practices that concern the robustness and reliability of results. Compared to findings from expert interviews conducted between 2010-13 (Hangel & Schickore 2017; Schickore & Hangel 2019), the practice of pre-registering hypotheses, designs, analysis methods, and the exploratory aspects of experimental studies on platforms like the Open Science Framework (OSF) has gained more traction. One of the consequences of this change is frontloading anticipated uncertainties by collaborative practices. Researchers now routinely address potential issues at various stages by aligning their experimental and writing-up practice with their pre-registered plans (as opposed to reconstructing an ex-post narrative). This change of practice not only mitigates publication bias but also influences how insights are disseminated across different mediums. Consequently, varying degrees of confidence in knowledge are now being communicated through diverse forms and outlets, which affects the tensions between accountable contribution and recognition.

Conclusion: This talk follows two aims: First, to present a study on reflections about tensions concerning evidence-checking processes and publication practices in the context of JDM. Second, to discuss the implications of publication bias as a social dilemma in JDM.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Communicating individual benefits promotes vaccination intention in the absence of strong social norms: A preregistered online experiment**

Aleksandra Lazic | Iris Zezelj

Faculty of Philosophy - University of Belgrade | Faculty of Philosophy - University of Belgrade

Rationale: Our previous experiments have shown that communicating a high vaccination uptake of 90% increases people's vaccination intention; however, from the public policy perspective, it is especially important to understand how to leverage the descriptive social norm as soon as the majority have been vaccinated but the uptake is still not high enough to eliminate the disease and protect everyone. In the present study, we tested what kind of messaging can motivate people to get vaccinated when 60% of others in their country have already done so. Based on a literature review and our study in which we elicited open-ended reports of reasons for people's vaccination choices, we hypothesized that, compared to baseline, (a) merely communicating the vaccination rate will increase an individual's vaccination intention (H1) but also that the following appeals communicated alongside the vaccination rate will be effective as well: (b) "protect your own health" (H2); (c) "protect the health of the people around you" (H3); and (d) "join the others to help stop the spread of the disease" (H4).

Methods: The study was preregistered and data was collected in an online survey advertised on Facebook in December 2023. Out of 1,303 participants who completed the study, N = 1,060 passed both attention checks and were included in the final sample. Participants were Serbian residents, 23.2% men, aged 18–77 (M = 47.8, SD = 12.9). All participants were first asked to imagine a fictitious scenario – with the symptoms of the disease and the vaccine-adverse events described as equally risky – and to assess their vaccination intention (baseline). They were then randomly assigned to one of the four experimental conditions (a–b, n = 265 each). Hypotheses were tested using repeated measures t-tests.

Results: While communicating the social norm alone was of no consequence (H1) – $t(264) = 0.08$, $p = .939$ – emphasizing the individual benefit of vaccination alongside the norm positively influenced vaccination intentions (H2) – $t(264) = 2.67$, $p = .008$, $d = 0.16$. Appealing to social (H3) and collective benefits (H4) alongside the norm did not change vaccination intentions ($t(264) = 1.18$, $p = .238$ and $t(264) = 1.36$, $p = .176$, respectively).

Conclusions: When the descriptive social norm is positive but weak (e.g. "60% of citizens are vaccinated") and therefore less likely to positively influence vaccination choices on its own, combining norms with individual benefits of self-protection may be a useful public communication strategy.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**The reputation consequences of punishment – Comparison of give-some and take-some games -**

Sakura Ono | Sotaro Aoki | Nobuyuki Takahashi

Hokkaido University | Hokkaido University

One of the most popular solutions to social dilemmas is punishment. However, the provision of punishment has a serious problem: the second-order dilemma. Since punishment is costly, it is better not to punish than to punish. Thus, free riding in punishment is a dominant choice, leading to mutual defection in the original SD. One of the solutions to this problem is provided by the acquisition of the positive reputation hypothesis, which states that punishers earn a good reputation by engaging in punishment. Then, the extra benefit from earning a good reputation may compensate for the cost of punishment. However, empirical evidence on the reputation consequences of punishment is mixed (Barclay, 2006; Kurzban et al., 2007; Raihani & Bshary, 2015). Hatano and Takahashi (2013) suggested that whether punishers earn a positive reputation depends on whether punishment is regarded as legitimate. Molenmaker et al. (2014) found that participants punished more in the take-some game than in the give-some game. Taken together, we expect the legitimacy of punishment and, eventually, the punishers' reputation to be higher in the take-some game than in the give-some game.

In order to examine whether punishers' reputation depends on the type of the game (take-some or give-some), we conducted a vignette experiment. Eighty-two undergraduate students participated in the experiment. They read a scenario describing an SD game (the between-subject factor - take-some or give-some). There was one defector and four cooperators. Then, they read a scenario describing two third parties, one of which is a punisher who engaged in costly punishment of the defector, and the other is a non-punisher who did not engage in punishment (the within-subject factor). Participants' responses were their impressions toward the punisher/non-punisher and their behavioral intention in the trust game (i.e., as a trustor how much they send to the trustee) if they were paired with the punisher/non-punisher.

A 2 x 2 mixed ANOVA revealed that the type of the game had no main or interaction effects on either impressions or the behavior of the trust game. On the other hand, the main effect of punishment was found, indicating that the non-punisher earned more favorable impressions than the punisher. Regarding the trust game, participants would send more toward the punisher than the non-punisher. These results suggest that to examine the acquisition of the positive reputation hypothesis, we need to consider the possibility that punishers' reputation is multi-faceted (Horita, 2010).

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Mapping perceptions of exploitativeness across situations and interactions**

Julia Teufel | Dr Maximilian Schmitt | Gerold Opoku | Dr Catia Pinto Teixeira | Dr Dr Hannes Rusch

Max Planck Institute CSL | Max Planck Institute CSL | Maastricht University | Maastricht University | Max Planck Institute CSL & Maastricht University

Rationale: Exploitative behaviour takes many different forms. Accordingly, different degrees of exploitativeness may be perceived across various interactions, including the workplace, romantic relationships and international relations. To better understand such perceptions across different relationships, we aim to determine those combinations of situational characteristics and behaviours that shape perceptions of exploitativeness.

Hypotheses: Our starting point is the meta-analysis of prosocial behaviour in economic games by Thielmann and colleagues (Psychol. Bull., 2020). Thielmann et al. hypothesize that exploitative interactions across games are characterized by extreme scores in two dimensions of interdependence between players. As such, these interactions entail (i) asymmetric dependence and power, with player A's outcome depending on player B's actions, as well as (ii) a conflict of interest between the players.

Testing Method: In a series of pre-registered studies, we plan to test this hypothesis based on real-life interactions elicited outside controlled lab environments. We will do this using new as well as existing data collected in experience sampling studies by Gerpott and colleagues (JPSP, 2018). Here, participants recounted a diverse range of situations they experienced in their daily life. The authors categorized these data into high and low scores on each of the five dimensions identified by interdependence theory, including the two dimensions used by Thielmann et al. in their description of situations that afford exploitation. Across multiple studies, (1) we will test if participants' perceptions of exploitativeness are in line with Thielmann et al.'s hypothesis that interactions with high asymmetric interdependence and power plus strong conflict of interest afford exploitation while others do not; (2) we will collect new data by asking participants to recount situations which they perceived as exploitative and score these along the dimensions of the Situational Interdependence Scale developed by Gerpott et al (2018). This provides a second and complementary test of the hypothesis by Thielmann et al., as these scores should follow the suggested pattern.

Intended Contribution: Since interactions outside the lab differ greatly from economic games, it is crucial to understand if they can indeed be described as suggested by Thielmann et al. (2020), or if additional dimensions are important. At ICSD, we intend to present a short talk or poster summarizing the results of our first empirical tests of this hypothesis as a starting point for future research on exploitative interactions in different relationships.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Gender bias in human collaboration with artificial intelligent agents**

Sepideh Bazazi | Jurgis Karpus | Taha Yasseri

Independent postdoctoral researcher, London, UK | LMU Munich, Germany | University College Dublin, Ireland

Thanks to advances in artificial intelligence, we may soon share roads with self-driving cars and work alongside automated software systems in pursuit of joint endeavours. It is important, therefore, to investigate whether human willingness to cooperate with others will extend to people's interactions with machines. While that is likely to vary across cultures, recent studies showed that people cooperate with artificial agents significantly less than they do with humans. One reason for that is people's greater willingness to exploit cooperative machines for selfish gain compared to their willingness to exploit cooperative humans.

A way to change this is to provide machines with human-like features. One understudied anthropomorphic feature of machines—yet, perfectly familiar to anyone who has used a voice activated GPS device—is machine gender. While there is some evidence that machine gender can influence people's behavioural dispositions, for example, willingness to donate money, how machine gender affects people's willingness to cooperate with them in mixed-motive "what's best-for-me-is-not-what's-best-for-us" settings is largely unknown. On the one hand, providing machines with gender may make people treat machines similarly to how they treat fellow humans and, hence, increase people's willingness to cooperate with them. On the other hand, human interaction with gendered machines may produce unwelcome side-effects, such as the reinforcement of gender stereotypes and the spillover of the impact of those stereotypes on human behaviour from human-human to human-machine interactions, or vice versa.

In this study we address two questions: 1) will human willingness to cooperate with gendered machines in mixed-motive settings be similar to their willingness to cooperate with fellow humans? 2) will existing gender biases in human-human cooperation extend to people's interactions with machines? To find out, we recruited human participants to interact with fellow humans and bots in the one-shot Prisoner's Dilemma game. To understand the motives behind people's willingness to cooperate with or defect against others, in addition to participants' own choices, we elicited their predictions about their co-player. We found clear differences in people's motivations and willingness to cooperate across genders of their interaction partners. We also found that gender biases present in human-human interactions extend to people's interactions with machines. In addition, people still treat machines somewhat differently from how they treat fellow humans.

Pre-registration: https://osf.io/38esk.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**The Role of Roles: The Impact of Roles on Behavior and Norms in a Public Good Game**

Rafael Nunes Teixeira | Sander Onderstal | J.W. Stoelhorst

University of Amsterdam | University of Amsterdam | University of Amsterdam

People assume various roles in their daily lives, such as fathers, mothers, teachers, and leaders, each entailing a distinct set of expectations and responsibilities. This study investigates the impact of social roles in a controlled laboratory setting, emphasizing its implications by analyzing the effects of designating a participant as a 'group leader' in a public goods game.

We explore the influence of social roles on both individual behavior and group dynamics by comparing groups with a leader and groups without a leader in two experiments. In our context, the leader is defined only by a label assigned randomly to one participant without any additional distinctions, focusing on the direct impact of social roles on behavioral changes and social norms.

In the first experiment, we analyze the behavioral changes associated with contribution levels in the public good game in groups with and without a participant labeled as a leader. Individuals labeled as leaders exhibit higher levels of contribution and reduced variance compared to their counterparts. Furthermore, participants are more likely to follow those assigned the leadership label than their counterparts, indicating an increase in influence on group dynamics.

In the second experiment, participants were asked to evaluate situations observed in the first half, indicating the expectations, norms, and social attributions associated with participants in groups with or without someone labeled a leader. We observe a shift in perception regarding the leader and the group. While people perceive the leader as someone who should contribute more, and they are expected to contribute more in the first round compared to counterparts, participants seem to expect that group dynamics prevail, describing similar empirical and normative expectations for group members contingent on similar contributions made by a leader or the member counterpart.

In summary, our experiments demonstrate that social roles lead to changes in behavior and social norms. Furthermore, these roles can be externally attributed and might be used to foster both societal and individual transformations.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Whose Norms to Follow? A Field Experiment on Both Sides of a Border**

Pascal Sulser | Urs Fischbacher | Irenaeus Wolff

Migros-Genossenschafts-Bund | University of Konstanz/TWI | TWI/Univ. of Konstanz

One of the well-known solutions to social-dilemma problems is the adherence to cooperative social norms. But how do people adapt to different norms and norm strengths across cultural boundaries? Is it more important where people come from or where they are? We investigate these questions in a field experiment. In particular, we focus on the public good not to litter and the associated norm. We study the behaviour and beliefs of Germans and Swiss across the German-Swiss border, by attaching small non-informative flyers to cars at parking garages in the border region and running two complementary online surveys. We find that behaviour depends more on where people are than where they come from. In fact, we not only see adaption to the local norm strength; in most cases, we see even over-adaptation. We use an illustrative model to guide our analysis and find support for the model in that littering decreases with the expected punishment, increases with the belief about the frequency of littering, and is lower in the country that cares more about the littering norm. If expectations were correct, they should only depend on where littering takes place. However, they also depend on where people come from, albeit not in the sense of non-residents `bringing their social-norms to another country'. Rather, it seems like contrast effects might be increasing perceived differences, and adaptation to the resulting beliefs then contributes to explaining the cases of `excessive adapation' we document.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Intuitive cooperation across group boundaries in a minimal group paradigm**

Kaede Maeda | Shigehito Tanida | Hirofumi Hashimoto

Rikkyo university | Taisyo university | Osaka Metropolitan University

Cooperation within and across group boundaries is important for the provision of public goods. However, in-group favoritism in cooperation, the tendency to cooperate more with ingroup members than with outgroup members, may limit contributions to public goods across group boundaries. The present study focuses on intuitive cooperation and the potential for intuition-based cooperation in dilemma-like situations across group boundaries, based on the group heuristic model (Yamagishi et al., 2007) and the model of intuitive cooperation (Rand et al., 2012). We hypothesize that intuitive cooperation in a one-shot prisoner's dilemma game (hereafter, PDG) is not possible across group boundaries and is limited among in-group members.

Three studies tested the above hypotheses. In Study 1, we used a minimal group paradigm, and observed participants' decision time in a one-shot PDG to compare the percentage of cooperation toward in-group and out-group members. Study 2 was a web-based experiment that manipulated the amount of time participants had to decide. Specifically, in the time constraint condition, participants were asked to decide whether to cooperate or defect in the one-shot PDG within 15 seconds; in the no time constraint condition, there was no such constraint as in Study 1. In Study 3, we attempted to replicate the study to confirm the robustness of the findings from Study 2 and also conducted an observation using an eye-tracking device to examine participants' decision-making process under time constraints.

We compared cooperation rates in the one-shot PDG in three studies. The results showed that in the no-time constraint situation, the rate of cooperation was higher only toward the in-group member. Interestingly, the results regarding the participants' decision time in the no-time constraint situation showed that the shorter the participants' decision time, the more cooperative the participants' choice was. In contrast, in the time-constrained situation, there were no group differences in cooperation rates, and the observed cooperation rates were relatively high regardless of the manipulated group condition. The results also suggest that when analyzing the gaze rate of the payoff matrix to examine participants' decision process, cooperators are more likely to look at the outcome of mutual cooperation.

Our hypothesis is not supported by the results. Rather, the results show that in the one-shot PDG, in-group favoritism occurs only in the absence of time constraints. Moreover, time constraints increase overall cooperation rates and thus eliminate the bounded group differences in the form of the minimal group paradigm.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**An Experimental Test of Risk Perceptions under a New Hurricane Classification System**

Jantsje M. Mol | Nadia Bloemendaal | Hans de Moel | Dianna Amasino | Jennifer M. Collins

University of Amsterdam | Vrije Universiteit Amsterdam | Vrije Universiteit Amsterdam | Tilburg University | University of South Florida

During a hurricane, it is vital that individuals receive communications that are easy to process and provide sufficient information to allow informed hurricane preparedness decisions and prevent loss of life. Without satisfactory and complete information, an individual is likely to miscalculate their personal risk or even potentially be moved to inaction. However, much recent research has shown an over-reliance on the currently-utilized Saffir-Simpson Hurricane Wind Scale (SSHWS) despite the fact that it only captures one aspect of a hurricane: the wind threat.

We study how the hurricane warning scale (traditional NOAA Saffir Simpson scale versus newly developed Tropical Cyclone Severity Scale) impacts intent to evacuate and understanding of hurricane severity. We use a between-subject design where participants are assigned to either the Saffir Simpson scale or the TCSS scale. We will collect data in a large-scale online experiment (N = 4000) to examine potential differences in comprehension, risk perception and anticipated evacuation and preparation decisions among citizens in U.S. coastal states under hurricane threat. We will test the hypotheses that the new scale increases understanding about the main hazard, increases evacuation intent for severe events, increases relevant precautionary measures (window protection for wind-driven storms, sandbags for rainfall-driven storms) but does not change worry.

Alternative hurricane hazard scales such as the TCSS may improve understanding of the general public, allowing for enhanced storm preparations and, ultimately, saving lives. Data collection is scheduled for April-May 2024. We look forward to discussing the results at the ICSD 2024 conference.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Intuitive cooperation in a one-shot prisoner's dilemma game with gain-loss frames: Revisiting the social exchange heuristic hypothesis**

Hirofumi Hashimoto | Yuka Mitsueda | Kaede Maeda

Osaka Metropolitan University | Osaka City University | Rikkyo University

Previous studies have shown that people tend to cooperate in a one-shot prisoner's dilemma game, despite the guarantee of anonymity. The argument proposed by Kiyonari et al. (2000) is a possible explanation for this phenomenon. In a one-shot prisoner's dilemma game, there is certainly no incentive to cooperate. Therefore, defection is a dominant choice. However, Kiyonari and colleagues insist that people subjectively bias the one-shot prisoner's dilemma game. Specifically, they argue that people perceive the prisoner's dilemma game as an "assurance game" because of an intuitive cognitive bias in processing information about social exchange. In the assurance game, there is no dominant choice. Defection results in an individually better outcome if the partner is also a defector. If the partner cooperates, however, cooperation produces an individually better outcome. With the subjective transformation, people intuitively perceive most mixed-motive incentive structures as ones in which mutual cooperation is personally more desirable - that is, produces personally better outcomes - than defection, provided that the partner also cooperates. It should be, therefore, reasonable for people playing a one-shot prisoner's dilemma game to recognize that mutual cooperation is most desirable according to this subjective transformation, i.e., the social exchange heuristic. This study aims to test Kiyonari et al.'s hypothesis, focusing on gain-loss frames and decision time.

Two studies were conducted to re-examine the social exchange hypotheses. In Study 1, we compared the effect of the gain and loss frames of the one-shot prisoner's dilemma game on self-reported satisfaction with the potential outcomes. In Study 2, we also examined the effect of the gain and loss frames of the one-shot prisoner's dilemma game and how participants' decision time related to their decision to cooperate or defect.

Our results supported the social exchange hypothesis proposed by Kiyonari and colleagues by showing that even in a one-shot prisoner's dilemma game, more than 60% of people cooperated regardless of the gain/loss frame, and most participants reported that the most desirable outcome was mutual cooperation. Interestingly, cooperators had significantly shorter decision times. The results suggest that the social exchange heuristic, which is an intuitive cognitive module, may be strongly activated in cooperators. However, the limitations of the traditional method of observing self-reported satisfaction were also pointed out, suggesting that this hypothesis needs to be tested in the future with the addition of other indicators such as decision time.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Do monetary incentives matter for measuring social preferences? A comparison of full-payment, random-payment, and hypothetical-choice incentive schemes in five economic games**

Lina Koppel | Amanda M. Lindkvist | Gustav Tinghög

Linköping University | Linköping University | Linköping University

The use of incentivized choice tasks is a long-standing tradition in economics, based on the belief that results from experiments can only be trusted if they involve real stakes. But does it matter whether participants are paid for all decisions in an experiment or a randomly selected decision? In this paper, we compare a full-payment incentive scheme (by which participants are paid for their decisions in all tasks) and a no-payment (hypothetical) incentive scheme (by which participants are paid for none of their decisions) to a random-payment incentive scheme (by which participants are paid for their decision in one randomly selected task), and investigate their effect on decision making in five standard economic games commonly used to measure social preferences: the Dictator Game, Ultimatum Game, Trust Game, Public Goods Game, and Prisoner's Dilemma. Results (n = 1,501) indicate that neither the full-payment incentive scheme nor the no-payment incentive scheme significantly predicted behavior above and beyond the random-payment incentive scheme. This finding held across all games included as well as a composite measure of prosocial behavior (B ranged from –1.934 to 0.533; all ps > .05). Our study suggests that the importance of real monetary incentives of the stake size typically used in experimental economics is overstated.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Prebunking Conspiracy Theories**

Matthieu Légeret | Jan Hausfeld | Jan Engelmann

University of Amsterdam | University of Amsterdam | University of Amsterdam

The spread of misinformation, such as conspiracy theories, is a growing concern in our digital era. Recent research showed that exposure to conspiracies led to a decrease in behavior necessary to contain Covid-19 pandemic. However, the effects of conspiracies on social behavior are sparse, and suggest a negative effect. Past research suggests that it may be more effective to counter misinformation by raising awareness about conspiracy theories before people are exposed to them (Lewandowsky & van der Linden, 2021; van der Linden, 2023).

Methods: In a preregistered experiment, we investigated how exposure to conspiracy theories affects beliefs and prosocial behaviours and how prebunking—a prevention intervention—can help negate the effects of conspiracy theory exposure. We randomly allocated 372 Prolific participants to one of three conditions: Control, Conspiracy, or Prebunking. Participants in the Control condition read an infographic about grass cutting, and then watched a video on the same topic. Participants in the Conspiracy condition also read an infographic about grass cutting but then watched a video promoting a conspiracy about the COVID-19 pandemic. Participants in the Prebunking condition read an infographic on how to identify conspiracy theories before watching the same video as participants in the Conspiracy condition. Subsequently, all participants took part in a series of economic games to measure social behaviours: a trust game (once as the trustor and once as the trustee), three dictator games with varying levels of initial inequality, and one charitable donation decision. We next assessed participants' beliefs in numerous conspiracy theories and their current emotional state. Finally, participants went through a debriefing on what they have seen in the experiment.

Results: Our preliminary analyses reveal that neither the exposure to a conspiracy theory nor the prebunking intervention significantly affected prosocial behaviours. However, we found that both conditions had a negative effect on participants' emotions (increasing negative emotions and decreasing positive ones). Furthermore, our results show that exposure to conspiracy theories, with and without the prebunking intervention, significantly increased the reported beliefs in related conspiracies. Finally, we exploratory analysis reveals that women and politically conservative participants tend to believe significantly more in conspiracy theories.

Conclusions: Our results indicate that prebunking may not be an efficient way to prevent the effects of exposure to conspiracy theories on emotions and beliefs, while social choices do not seem to be affected by conspiracies.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Kill some to save many ? when I feel depressed, I'll think about it.**

Kévin Bague | Maxime Bourlier | Cassandra Leroux | Jean Baratgin

Université Paris 8 | Université Paris 8 | Université Paris 8 | Université Paris 8

Research on moral reasoning has used sacrificial dilemmas to explore the conflict existing between two moral positions: Utilitarianism which is the maximization of happiness for the most and Deontologism which is when we follow a predetermined moral code. In sacrificial dilemmas, we are given a choice between saving multiple lives by sacrificing one (i.e. the Utilitarian choice) or refusing to make this sacrifice (i.e. the Deontic choice). According to the Dual Process Theory of moral reasoning, there are two types of processes. The intuitive process, which is fast and automatic, would lead to Deontic responses. On the other hand, the deliberative process is slower, conscious and requires cognitive resources. Since Utilitarian responses have been shown to usually take more time to be given, they have been associated with the latter process. However, exclusivity of the two moral responses to each process is put into question by a new model of the dual process theory. In contexts where the kill/save ratio is render extreme such a killing one person to save five hundred, Utilitarian responses come intuitively to mind as their low response times suggest. In this new hybrid model, the intuitive process would generate both Utilitarian and Deontic response with various degrees. We would only switch to the deliberative process when a conflict that makes us unable to answer occurs between those intuitions. The deliberative process' function would be to resolve this conflict. In this study, we submitted 120 participants to multiple kill/save ratios ranging from 1 for 1 to 100 for 5000 in three different scenarios and assessed their levels of depressive symptoms to explore the transition from the intuitive process to the deliberative process and back. Participants with high depressive symptoms scores give more utilitarian answers even for the lower ratios but also have higher response times on average.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Protected, but at what cost? Investigating the potential side-effects of inoculating against misinformation**

Teodora Spiridonova | Olga Stavrova | Ilja van Beest

Tilburg University | Lübeck University | Tilburg University

The rise of misinformation in recent decades has had many negative consequences for individuals and societies alike. One promising strategy for countering such misinformation is “prebunking”, or inoculation – exposing people to different misinformation tactics, to prevent them from falling for misinformation in the future. Contrary to “debunking”, which relies on correcting misinformation after the fact, inoculation has been described as immune to side-effects. We aimed to test this notion – specifically, we investigated the possibility that inoculation lowers the perceived accuracy of real, as well as fake news. Furthermore, based on research showing that even mere exposure to discussions about misinformation can make people more cynical towards politicians and the media, we hypothesized that inoculation would increase both news-related and general cynicism. To maximize ecological validity and account for stimulus bias, we tested our hypotheses using a large-scale dataset of existing real and fake news headlines (n = 120), and used a random stimuli approach. We exposed 385 participants to either an inoculation intervention or a control text. Participants then rated the perceived accuracy of ten headlines (half of which contained misinformation) and responded to measures of news-related and general cynicism. Results of a multilevel regression showed that participants exposed to the inoculation (vs. control) condition rated fake news headlines as less accurate compared to real news headlines. Furthermore, there was no effect of condition on either general cynicism or news media cynicism, indicating that inoculation interventions do not render people more cynical towards the news media or other people. Our findings thus provide further evidence for the robustness of inoculation interventions against side-effects and contribute to the literature on perceptions of fake news.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Roll for the future: Introducing a novel climate change game**

Patricia Kanngiesser | Jan K. Woike

University of Plymouth | University of Plymouth

Human-made climate change is one of the most pressing challenges of our times, threatening lives and livelihoods. Mitigating the impacts of global heating and reducing global carbon emissions requires global cooperation. However, climate change and its underlying dynamics are a complex system with non-linear growth dynamics, delayed impacts, and a conflict between self-interests and societal benefits (Weber & Stern, 2011). Games allow people to experience these dynamics in the safety of a simulated environment. Simulated game environments have been used to study climate change related behaviors in the lab (e.g., Milinski et al., 2008; Tavoni et al., 2011) and, at the same time, there is an increasing number of climate change themed board games and simulations (e.g., Kwok, 2019). Here we present a new incentivized climate change game for experimental studies both offline and online. The game is played with gameboards and dice in different colours. It includes exponential dynamics and involves decisions between short-term point maximization (risking adverse consequences) and long-term investments that mitigate exponentially increasing adverse consequences. The game can be played in a neutral or a climate-change frame, and as a solo or group version, adding a social dilemma to the game (i.e. potential adverse consequences are shared by everyone in the group). In this talk, I will present the game structure as well as data (currently being collected offline) on solo game play behaviour in a neutrally-framed version of the game and its relation to different social and cognitive factors (e.g., numeracy, risk, cognitive reflection, social value orientation, etc.).

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Together for a common cause? Cooperative tendencies in transdisciplinary research groups aimed at solving water quality and quantity issues in Eastern North Carolina**

Kyra Selina Hagge | Delnaz Amroliwalla | Marcia R. Hale | Poonam Arora | Stephen Moysey

Department of Coastal Studies, East Carolina University | School of Business, Quinnipiac University | Department of Peace & Conflict Studies, UNC Greensboro | School of Business, Quinnipiac University | Department of Geology, East Carolina University

Complex challenges around water management, governance, and financing demand a trans-disciplinary approach to research, which integrates multiple disciplinary lenses, includes the grounded knowledge of community expertise, and has the potential to be transformative in its impact. Because time and effort are limited resources for researchers, they face a social dilemma (Dawes & Messick, 2000): should they invest their time in furthering their own research, or transdisciplinary efforts? Additionally, publishing in disciplinary-specific journals may receive greater recognition and funding. We address the trade-offs faced by researchers in transdisciplinary projects as a nested social dilemma and help to understand the challenges intrinsic to engaging in transformative research. Does individual Social Value Orientation (SVO) impact people's willingness to share resources within their own disciplinary group (in-group) and the larger project group (out-group)? How do researchers define the relationship between their in- and out-group and transcend the lines in between?

Our NSF Coastlines and People (CoPe)-funded project includes more than 30 scientists in five working groups collaborating with community partners to address the environmental justice dimensions of water quality and quantity issues in Eastern North Carolina. Grant colleagues responded to a questionnaire with the SVO and a modified version of the Intergroup Parochial and Universal Cooperation (IPUC) game at two team events, providing a baseline measure for individual cooperation.

Data collected in October 2023 (N=17) and November 2023 (N=11) together provide baseline SVO measurements for 28 individuals working on the CoPe research grant (N=22 pro-social, N=6 individualistic).

Corresponding with findings by Aaldering & Böhm (2020), and exploring the modified IPUC data, we observe pro-self team members were more likely to keep money for themselves while the pro-social oriented researchers contributed more to the universal pool, which benefits the in- and out-group equally. Physical science team members contributed an average of 2 additional coins to each of the weak parochial and egoism pools. Conversely, the average contributions by the social science team members were 7.47 to the universal pool, and 1.8 to the weak parochial pool with less than 1 coin contributed to the egoism pool.

We are currently collecting data on whether trade-offs in transdisciplinary research are more likely to be perceived as an assurance -, maximized payoffs -, or chicken game, or prisoner's dilemma. Building upon our results, we will design interventions to increase cooperative behavior and team connectivity that can result in higher levels of transdisciplinary research.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**The development of social prediction during adolescence**

Jaime Vigil Escalera | Ili Ma

Leiden University | Leiden University

Social life undergoes important changes during adolescence, where relationships with peers begin to matter more and new social challenges arise. Successful adaptation to these changes heavily depends on the ongoing development of social skills. An important skill is social learning. Specifically, learning about others' motivations is necessary to predict their behavior in new contexts. In this preregistered study, we investigate the development of learning others' social preferences during adolescence. Participants (n = 61 in the age range 10 to 20 years old) completed a behavioral task in which they were asked to predict the cooperation or defect choices of four other players, who each played a series of behavioral economic games with unique payoff matrices. Each player made choices according to their own motivation (the four motivations were: greed, risk aversion, and their inverse, each corresponded with a unique player). Due to the varying payoff matrices, each player cooperated on 50% of the trials. In order to perform above chance, participants therefore needed to learn the underlying motivation that governed the player's choices. Preliminary results show higher accuracy when predicting behavior motivated by greed and risk aversion than inverse greed and inverse risk aversion. Additionally, our analysis suggests that while prediction accuracy increases with age, this effect differs between commonly used strategies (i.e., greed and risk-aversion) and their artificial counterparts (i.e., inverse greed and risk-aversion), with the latter improving more with age. Currently more data is being collected to confirm these results (target n=150). Moreover, these data will be analyzed using feature-based reinforcement learning models to examine age-related changes in this learning process. With this study, we aim to shed light on how adolescents learn to predict how others will behave, which helps them navigate the growing complexity of their social lives.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Correlates of Social Preferences**

Adam W. Stivers | Carson Oliver | Zach Nagy

Gonzaga University | Whitworth University | Gonzaga University

Rationale: The purpose of this research was to examine relationships between Social Value Orientation (SVO), Social Mindfulness, and a variety of individual difference measures.

Method: 380 undergraduate students completed a survey with a battery of short measures that included the Slider Measure of Social Value Orientation (SVO), the Social Mindfulness Measure (SOMI), the HEXACO-60 personality inventory, the 80-item Zuckerman-Kuhlman-Aluja Personality Questionnaire (ZKA-PQ/SF), the Dirty Dozen Measure of the Dark Triad, a measure of Generalized Trust, and a variety of demographic questions.

Results: In relation to the HEXACO-60 personality inventory, SVO was found to have correlations with Honesty-Humility ($r = .29$, $p < .001$), Conscientiousness ($r = .12$, $p = .023$), and Openness ($r = .13$, $p = .01$). Relationships with Emotionality, eXtroversion, and Agreeableness were all non-significant. SOMI was found to have a positive correlation with Honesty-Humility ($r = .17$, $p < .001$), while relationships with all other HEXACO dimensions were non-significant.

Conclusions: The relationships between SVO and the dimensions of the HEXACO-60 inventory are generally consistent with prior literature. For SOMI, the relationship with Honesty-Humility replicates prior research, but a relationship with Agreeableness was also expected. Relationships with additionally measures will be reported when data analysis is complete.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**The evaluation of third-party punishment depending on its type, severity, and interpersonal hierarchy**

Olivia Seubert | Anne Böckler-Raettig

University of Würzburg | University of Würzburg

Third-party punishment is regarded as an altruistic act, as it helps foster cooperation, trust, and fairness in social groups. However, punishment can also indicate anger and potential threat, and recent studies highlighted the ambiguity of punishment as a signal of cooperative intent. We aimed to go beyond the highly abstract and decontextualized settings typically employed in economic games to better understand how punishment and punishers are perceived in interdependent real-world situations.

Therefore, we created and validated 24 written everyday-life-scenarios and, in each of them, manipulated the type of transgression/punishment between perpetrator and victim/punisher and perpetrator (property-oriented, corporal, psychological; Experiment 1), the severity of punishment (weak, strong; Experiment 2) and the hierarchy between punisher and perpetrator (punisher equal or higher in rank; Experiment 3). After each scenario, participants rated the punishment's adequacy and the punisher's warmth, competence, and suitability as an interaction partner, whether as a friend or team leader. Results indicated the preference of punishments aligned with the type of transgression, weaker punishments, and punishers equal in rank to the perpetrator. Across all experiments, third parties engaging in psychological punishment were preferred and preferentially chosen as potential interaction partners. Our findings support the notion that punishment is a delicate issue, and reveal interpersonal and contextual factors that contribute to its evaluation as a useful strategy.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**The dual influence of inetrpersonal bonds on the cognitive processes of disobediance**

Nicolas Coucke | Daniele Marinazzo | Salvatore Lo Bue | Emilie Caspar

Ghent University

Most humans have a strong tendency to obey the orders of superiors, even when these orders may cause harm to others. Previous studies pointed out that prosocial disobedience to such orders might be supported by empathy towards a potential victim. Meanwhile, a close relationship with a superior may increase the tendency to obey orders. We hypothesize that both one's relationship with a potential victim and one's relationship with a commander might modulate the ease with which individuals obey or disobey immoral orders. To explore this, we designed an experimental paradigm where 3 participants were each assigned a fixed role of either victim, agent, or commander. Initially, the agent engaged in a joint drawing task with both the victim and the commander to establish interpersonal relationships. In a subsequent phase, the agent was instructed to either obey or disobey the commander's orders to administer moderately painful shocks to the victim. Simultaneous EEG recordings of the 3 participants were made during the entire experiment. Preliminary results indicate that cognitive conflict, as indicated by mid-frontal theta activity, during antisocial obedience is negatively correlated with the subjective strength of the agent-commander relationship. These findings shed light on how complex social situations influence cognitive processes related to obedience and disobedience.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Trust-based information filtering can form polarising group identities**

Fredrik Jansson | Anandi Hattiangadi | Magnus Enquist

Mälardalen University | Institute for Futures Studies | Stockholm University

Rationale: This study examines how trust-based influence leads to cultural polarisation and identity formation. In polarised societies, individuals often align beliefs and actions with their group, creating entrenched viewpoints and resistance to compromise. This can exacerbate social dilemmas, as polarised groups are less likely to cooperate or find mutually beneficial solutions. Polarisation has been explained by irrational learning mechanisms, similarity bias or informational segregation. We have developed a model to study how groups form and become associated with sets of beliefs that are not necessarily interrelated, leading to the emergence of social identities. Contrasting to previous models, we assume full network connectivity and conditional openness to learning from others with consistent beliefs.

Methods: We developed a mathematical model where agents can hold beliefs in propositions or counter-propositions, for example, P can be a belief in human-induced climate change and Q that vaccines are safe and effective, and not-P and not-Q their opposites. Agents trust those who hold beliefs not inconsistent with their own (sender filtering). We also compare this to an alternative model (belief filtering) where agents consider the belief in question rather than the sender, and we also studied the impact of external signals and empirical feedback. We analysed the tractable part of the model and ran simulations to include more potential beliefs.

Results: Sender filtering can cluster the population into strongly connected belief packages. Initially unrelated beliefs can become strongly correlated, with the largest factions typically polarised at the extremes. For example, if one group believes P and Q, on climate change and vaccines, then the other typically believes both not-P and not-Q. The model shows that these belief clusters function as identity signals, where adherence to one belief increases the likelihood of adopting other beliefs within the same cluster. This contrasts with belief filtering, where polarisation occurs only if beliefs are intrinsically connected.

Conclusions: Sender filtering can lead to extreme polarisation and formation of belief packages that act as identity signals within groups. These results can help in understanding how cultural underpinnings influence social dilemmas, or create dilemmas through obstructing coordination, particularly in terms of belief systems shaping group identity and intergroup relations. Meanwhile, the result that belief filtering leads to less polarisation hints at potential countermeasures. Finally, the results suggest that digital technologies, through algorithms like collaborative filtering, could exacerbate polarisation and identity formation, posing new challenges for managing social dilemmas in a digital age.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Aversive medical treatments signal a need for support**

Mícheál de Barra | Daniel Cownden | Fredrik Jansson

Brunel University, London | None | Mälardalen University

Objective: Ineffective, aversive and harmful medical treatments are surprisingly common cross-culturally, historically and today. Meanwhile, humans are often incapacitated by illness and injury, and are unusually dependent on care from others during convalescence. Over life, we thus stand to gain from giving care to needy and receiving care when in need. However, such caregiving is vulnerable to exploitation via illness deception, whereby people feign or exaggerate illness in order to gain access to care. The question of giving and receiving care is thus a social dilemma, and we suggest that aversive treatment may have persisted as part of a solution.

Methods: We formulate an evolutionary game theoretical model where individuals can be healthy or sick, and have a strategy of whether to ask for help, at a cost to the helper (that causes reduction in the spread of the sender's behaviour) and a benefit (increasing spread of the behaviour) to the recipient, if provided, and whether to provide help when asked. Specifically, we are interested in the conditions under which providing help is a stable strategy. Our main question is whether the range of conditions where helping is evolutionarily viable can be increased through the introduction of aversive medicine.

Results: There is a broad range of conditions where the possibility of deception undermines caregiving. However, it is possible for caregiving to become an ESS, via a judicious choice of the degree of aversiveness of the treatment. To the extent that illness deception undermines caregiving, aversive medicine can help prevent this erosion. That is to say, aversive medicine plays a more important role when illness deception is common.

Conclusions: Our model demonstrates that aversive treatments can counter-intuitively increase the range of conditions where caregiving is evolutionarily viable, because only individuals who stand to gain substantially from care will accept the treatment. Thus, contemporary and historical "ineffective" treatments may be solutions to the social dilemma of allocating care to people whose true need is difficult to discern.

Poster session | Poster room

Tuesday July 2, 2024 17:00 - 18:30

**Cooperation in Nested Social Dilemmas: The Role of Pooled Punishment**

Misato Inaba | Yoko Kitakaji

Kindai University | Hiroshima University

Cooperation across group boundaries is vital for addressing multiple group-related challenges, including climate change. A nested social dilemma framework involving non-cooperation, local cooperation, and global cooperation is essential for cross-group cooperation. Kitakaji and Inaba (in prep.) demonstrated that peer punishment, where the punisher choose whim to punish, fosters global cooperation in some groups and local cooperation in others within nested social dilemmas. This results from communicating cooperation expectations through punishment, establishing group-specific cooperation norms. However, punished individuals may not understand the behaviors leading to punishment and how to change their behavior accordingly. This study addresses two questions: 1) What behaviors do players perceive as deviations from the cooperation norm? 2) Does clarifying the punishing intent lead to global cooperation? In a 20-trial nested social dilemma experiment, 108 participants were divided into six-person groups, each comprising two three-person local groups. Participants were tasked with choosing to cooperate globally, locally, or abstain from cooperation. Experimental conditions included a control group (no punishment) and a pooled punishment group. In the pooled-punishment condition, participants exhibiting the lowest levels of global or local cooperation within the society were subjected to punishment utilizing two types of punishment funds for global and local cooperation. A sample size of 98 was determined for detecting effects at the f = 0.1 level (? = 0.05 and d = 0.50) for primary and interaction terms through a sensitivity power analysis using G*Power. The target sample size was set at 108 to accommodate a six-person group game under two conditions. This study's design, hypotheses, and analysis plan were preregistered. We conducted a linear mixed model using log-ratio analysis, with condition between participants and cooperation type (local or global referencing non-cooperation) within participants as fixed effects. Random effects included a six-person group, three-person group, and participant. The analysis revealed higher levels of local and global cooperation in the pooled punishment condition compared to the control group. Non-cooperation decreased, with participants exhibiting equal levels of cooperation at both global and local levels. Additionally, the group applied the same degree of punishment for less global cooperation as for less local cooperation. Explicit punishment for cooperative norms resulted in decreased non-cooperation while maintaining cross-group and within-group cooperation in most instances. However, universal attainment of global cooperation, which maximizes social benefit, was not achieved. Individuals might only perceive non-cooperation as punishable, suggesting no distinct preference or consequences for local or global cooperation.