2nd Matterhorn Symposium on Behavior, Institutions, and Cooperation
September 18 – 20, 2025

UniDistance Suisse
Schinerstrasse 18
3900 Brig, Switzerland
https://unidistance.ch/matterhornsymposium

Book of Abstracts
(grouped by session)

Note: In the parallel sessions, each presenter has 30 minutes and may decide how to divide this time between presentation and discussion. The last presenter in each session is responsible for keeping time.

13:45 – 15:15   Session 1L (Seminar Room B18.003)

Interpersonal Trust During Times of Political Unrest
Maltz Amnon

Do extreme political events affect interpersonal trust, and does their impact depend on individuals' perceptions of these events? We test these questions using the controversial judiciary reform introduced in Israel in January 2023 by the most right wing government in the country's history. This reform was, and still is, highly controversial among the Israeli society, and viewed by many as threatening the very democratic nature of the state. Using an online incentivized survey administered to a large representative sample of the Israeli adult population, we compare interpersonal trust before the elections (October 2022) and during the political turmoil that has spread throughout the country following the advancement of the reform (March 2023). We report two main findings. First, interpersonal trust dropped across the board. Second, the most pronounced decline was observed among those who hold neutral attitudes towards the reform. Prior to the political events, they had higher levels of interpersonal trust than the ideological groups, but following the political events, their trust levels became indistinguishable from the other groups. We propose a model that links trust to ideological distance to account for the patterns found in the data.

Revisiting Situational Strength: Do Strong Situations Restrict Variance in Behaviors?
Ranran Li

The idea that strong situations—those involving high stakes, clear expectations, and strict constraints—restrict variance in behavior has long been treated as a maxim in psychology. Yet empirical support for this "restricted variance hypothesis" has remained inconclusive. To address this gap, we conducted a preregistered meta-analysis (k = 301, N = 25,670) in the context of cooperative behavior observed in standard social dilemma paradigms.
Our findings provide robust support for the hypothesis: Strong, compared to weak, situations—both theoretically defined and empirically validated through perception ratings—consistently restricted behavioral variance. Specifically, several structural features of social dilemmas—including the presence of punishment, low (vs. high) anonymity, exogenous (vs. endogenous) institutional rules, and low (vs. high) uncertainty—were associated with reduced variability in cooperation. Additional ratings of perceived situational strength in a subset of studies (k = 138, n_studies = 41) further confirmed that stronger perceived situations correlated with less behavioral variance.
These findings not only validate a core tenet of the situational strength framework but also shed light on the psychological mechanisms through which strong situations exert their restricted variance effects. We propose that reinforcement learning (via outcome-based incentives or punishments), social norms and reputation concerns (via observability and institutional enforcement), and cognitive dissonance reduction (via clarity and consistency of expectations) jointly contribute to the restriction of behavioral variability. The findings also have important theoretical implications for advancing research on person–situation interactions.
Beyond their theoretical contributions, our results have practical relevance for behavioral economists and policy designers aiming to foster predictable, cooperative behavior in domains such as public goods provision, regulatory compliance, and collective action.

War Causes Religiosity: Gravestone Evidence from the Vietnam Draft Lottery
Wladislaw Mill

Does war make people more religious? Causal evidence on this important question has long been wanting. To provide that evidence, we exploit the Vietnam Draft Lottery ‚Äì a natural experiment that drafted male U.S. citizens into military service during the Vietnam War. We measure religiosity via religious imagery on web-scraped photographs of hundreds of thousands of gravestones of deceased

U.S. Americans using a tailor-made convolutional neural network. Our analysis provides compelling and robust evidence that war indeed causes higher religiosity: people who were randomly drafted to go to war are at least 20\% more likely to have religious imagery on their gravestones. This effect holds up against multiple robustness and placebo tests, sets in almost immediately, persists even after 50 years, and generalizes across space and societal strata.

Session 1R (Seminar Room B18.005)

Zero-Sum Thinking and Support for Redistribution Across Borders
Anca Balietti

Rising nationalist and protectionist rhetoric increasingly frames global economic relations in zero-sum terms. This paper examines how such beliefs shape support for international redistribution. In a pre-registered two-by-two experiment with a representative sample of 2,116 UK adults, we manipulate (i) whether participants adopt a zero-sum or positive-sum worldview, and (ii) whether they are informed of their income rank in the global distribution. Participants then decide whether to donate to an international anti-poverty organization. Zero-sum priming significantly reduces support for global redistribution. However, this effect is fully offset when participants are informed of their relatively advantaged position in the global income distribution. A simple theoretical model accounts for this interaction by linking zero-sum beliefs and redistribution preferences to awareness of global income position.

Redistribution, Moral Hazard, and Voting by Feet: An Experiment
Alisa Frey

This paper presents a laboratory experiment investigating how individuals choose between different redistribution levels, incorporating moral hazard and voting by feet dynamics. While survey data suggests many people have preferences for redistribution, actual voting behavior, even when it would be economically beneficial, can be counterintuitive. This study goes "one step forward" by examining actual choices when individuals can select between fixed redistribution rates, mimicking "voting by feet" across different regimes. Two key challenges are explored: system competition potentially pressuring high redistributive systems and inefficiencies arising from moral hazard where citizens work less under higher taxes. The experiment lets subjects choose between two purely redistributive tax regimes: a low tax (15%) implying little redistribution and a high tax (85%) leading to substantially more equal incomes. After choosing, subjects earn income through a real effort task (decoding puzzles). The process is repeated six times to allow for learning, fatigue, and the potential unraveling of tax choices. The study addresses several research questions, including what tax rate subjects choose over time (distinguishing initial choice under a veil of ignorance from later, potentially opportunistic choices), the effort differences between tax regimes, and how these relate to self-selection. Control treatments with exogenous tax rates (Tax Random) and random income (Income Random) are used to help identify the moral hazard effect and subjects' attitudes toward income uncertainty. Key findings reveal that subjects' initial tax choices behind a veil of ignorance are heterogeneous, close to a 50:50 split between high and low taxes, consistent with other studies. However, underconfident subjects are more likely to choose the high tax. Over time, there is a significant decrease in the proportion of subjects choosing the high tax regime, supporting the hypothesis of gradual unraveling and voting by feet. Subjects tend to choose the tax rate opportunistically based on their previous round's status: Net receivers are more likely to choose the high tax, while net payers are more likely to choose the low tax. The experiment confirms a significant moral hazard effect of taxation, with subjects solving fewer puzzles at the high tax rate. This effect is largely driven by above-median performers. A substantial selection effect is also observed, where high-effort subjects migrate to the low tax regime over time. More experienced experiment participants are less likely to choose the high tax rate in subsequent rounds. In conclusion, the study shows how redistributive preferences translate into actual choices and how awareness of individual and relative performance leads to unraveling and self-selection, ultimately reducing support for redistribution over time. These findings have relevance for real-world issues like location choice in the context of tax competition and mobility.

Inequality, Life Expectancy, and the Alienation Effect: Insights from a Real-Effort Experiment on the Intragenerational Redistribution Puzzle
Tim Krieger

This paper addresses the "intragenerational redistribution puzzle" in public pension systems: Although rising income inequality leads to increasing differences in life expectancy – which would normatively call for more redistribution in favor of low earners – the actual progressivity of many OECD pension systems has declined since the 1980s. We develop a behavioral economics model that explains this paradox through an alienation effect: Increasing mortality risks among low-income cohort members reduce the willingness of decision-makers to redistribute, especially when they perceive efficiency losses (the "leaky bucket" problem).

The model combines an individualistic social welfare function with three psychological mechanisms:
(i) a trade-off between risk/inequality aversion and efficiency,
(ii) self-identification with future income positions (the "equiprobability rule" vs. wishful thinking about upward mobility), and
(iii) the weighting of distributive neutrality.
These mechanisms imply that asymmetric lifetime risks do not automatically lead to more redistribution – in fact, under reduced system efficiency, they can result in less.
To empirically test the model, we conducted a pre-registered online lab experiment (n = 144), where participants acted as "impartial stakeholders" behind a veil of ignorance to set the redistribution parameter of a stylized pay-as-you-go system. In three between-subjects treatments, we varied (a) symmetric vs. asymmetric mortality risks and (b) the presence or compensation of efficiency loss. Four within-subject rounds varied (i) contribution variance and (ii) benefit level. Participants earned income positions through a real-effort slider task, behaviorally anchoring institutional entitlements.
Our key findings are:
First, we observe a significant alienation effect: Under asymmetric mortality risk with efficiency loss, median redistribution drops by 12 percentage points compared to the symmetric benchmark. This drop vanishes

when efficiency loss is fully compensated – distributive neutrality alone does not trigger more redistribution. Second, the effect remains robust across benefit levels and contribution variances, suggesting a general behavioral mechanism.

Third, individual preferences and beliefs explain behavioral heterogeneity: Those with high risk tolerance, strong belief in social mobility, just-world beliefs, or victim-blaming tendencies reduce redistribution especially strongly under efficiency loss.

The study links behavioral traits (social preferences, wishful thinking, risk attitudes) with institutional design parameters (pension formulas, efficiency incentives), offering new insights into cooperative solutions for societal challenges. It shows that cooperation in solidary old-age provision is not only shaped by formal rules, but critically by perceived efficiency and fairness. Reforms that highlight or render efficiency losses salient may unintentionally undermine solidarity. Yet, if losses are compensated or minimized via smart incentives, high redistribution and social cohesion can be preserved even amid growing lifetime inequality. This contributes to the Matterhorn symposium's core theme – the interplay of behavior, institutions, and cooperation – and shows how behavioral economics can help build resilient social policy institutions.

15:45 – 17:15 Session 2L (Seminar Room B18.003)

Procedural Decision-Making in the Face of Complexity
Gonzalo Arrieta

A large body of work documents that complexity affects individuals' choices, but the literature has remained mostly agnostic about why. We provide direct evidence that individuals use fundamentally different choice processes for complex and simple decisions. We hypothesize that individuals resort to "procedures"— cognitively simpler choice processes that we characterize as being easier to describe to another person— as the complexity of the decision environment increases. We test our hypothesis using two experiments, one with choices over lotteries and one with choices over charities. We exogenously vary the complexity of the decision environment and measure the describability of choice processes by how well another individual can replicate the decision-maker's choices given the decision-maker's description of how they chose. We find strong support for our hypothesis: Both of our experiments show that individuals' choice processes are more describable in complex choice environments, which we interpret as evidence that decision-making becomes more procedural as complexity increases. We show that procedural decision-makers choose more consistently and exhibit fewer dominance violations, though we remain agnostic about the causal effect of procedures on decision quality. Additional secondary evidence suggests that procedural decision-making is a choice simplification that reduces the cognitive costs of decision-making.

Compressed Beliefs
Christian Zihlmann

This paper provides compelling evidence for belief compression and proposes a method to correct it. A controlled lab experiment, a natural field experiment with sports bettors, and a large-scale survey experiment on official inflation expectations conducted by the German Bundesbank all point in the same direction: probabilistic beliefs are compressed toward a cognitive default, which assigns equal probability to all partitions the state space was divided into—the ignorance prior (e.g., 50-50 in the binary case). This implies that probabilistic beliefs, elicited in a behaviorally compatible way (Danz, Vesterlund and Wilson, 2022), are malleable: they depend causally on the researcher's (implicit or explicit) choice of state space partitioning. Moreover, belief compression leads to biased means, too little variance, and attenuation bias. I present a method to predict the magnitude of compression in elicited beliefs, and to reverse it by inferring decompressed beliefs. The method is simple to implement, even in field settings. I demonstrate its value across the three datasets. Compared to elicited beliefs, inferred beliefs are better at predicting induced beliefs, actual realizations of events, and individuals' own betting behavior. To illustrate broader implications, I replicate the gender gap in confidence documented in Niederle and Vesterlund (2007) and Exley and Nielsen (2024). I show that the magnitude of the gap is malleable, as it is endogenous to the choice of state space partitioning. Moreover, we underestimate the gap when using elicited beliefs due to attenuation. The confidence gap grows in both magnitude and statistical significance when using the decompressed inferred beliefs. Eliciting probabilistic beliefs is no easy task. A central focus of the literature has been on designing incentive-compatible mechanisms for eliciting beliefs, first in a theoretical sense, and more recently in a behavioral sense (Danz, Vesterlund and Wilson, 2024). This paper proposes expanding that focus to study the behavioral compatibility of the elicitation task itself—here, the partitioning of the state space.

The Role of Noisy Coding and Serial Dependence in the Description-Experience Gap
Gökhan Aydogan

The Description-Experience Gap (Barron & Erev, 2003; Hertwig et al., 2004) highlights differences in economic decisions when outcomes and probabilities are explicitly described versus learned through experience. Typically, small probabilities are overweighted in description-based decisions compared to experience-based decisions, reflecting distinct risk tolerances. Although previous models have considered memory biases, noisy probability encoding, and recency or primacy effects, the underlying perceptual distortions remain incompletely understood.

Building upon noisy coding models (Petzschner et al., 2015), we propose a Bayesian perceptual framework incorporating serial dependence—the influence of sequentially encountered outcomes and probabilities on subsequent perceptions. Specifically, we hypothesize that serial dependence induces predictable probability distortions, differentiating decisions made from described versus experienced information. In experienced-based decisions, initial extreme outcomes bias perceived probabilities toward these extremes (primacy effect). Conversely, description-based probabilities tend to regress toward the provided distribution's mean. We tested these predictions using an incentivized binary lottery experiment with 166 participants. Each participant completed 100 trials under both description-based and experience-based conditions. Lotteries comprised either a low-probability, high-reward "risky" option or a high-probability, low-reward "safer" option. In experience trials, participants observed 10 outcome samples per lottery before making a choice, with outcomes matched to description-based probabilities. Results strongly supported our serial dependence hypothesis. Description-based decisions exhibited systematic overweighting of small probabilities and underweighting of large probabilities, consistent with

known decision-making biases. Conversely, experience-based decisions demonstrated the opposite distortion, reflecting an underweighting of small probabilities and overweighting of large probabilities driven by sequential outcome sampling. To quantify these perceptual distortions, we implemented a Bayesian perceptual model incorporating serial dependence and compared it against canonical models (e.g., Prospect Theory). Model comparison using Deviance Information Criterion (DIC) confirmed that our perceptual model provided the superior fit, successfully capturing both qualitative and quantitative aspects of the observed behavior.
These findings indicate that serial dependence significantly shapes subjective probability perception, underpinning the Description-Experience Gap. Our model explains how sequentially processed information induces distinct probability distortions depending on decision context. Understanding this perceptual mechanism is crucial for developing interventions and financial regulations targeting systematic risk perception errors. Future research should explore serial dependence effects in real-world financial decisions and contexts involving significant stakes, such as investment and insurance.

Session 2R (Seminar Room B18.005)

Charismatic Leadership: An Antidote to the Pitfalls of Incentives?
Christian Zehnder

Past research has established that charismatic leadership tactics can be a powerful motivator. In some settings, the increase in work output induced by a charismatic speech is comparable in size to the positive effect of high-powered financial incentives. But what about settings in which incentives backfire? In a between-subject laboratory experiment, we set up a real-effort work environment in which participants can execute a task in two ways: they can either "work hard" so that each produced unit creates a sizable benefit for the principal, or they can "take shortcuts", which takes much less effort but also substantially reduces the benefit of a produced unit for the principal. When compensation is a fixed wage and the motivation speech is ``standard'', we observe that participants mostly focus on the socially optimal, hard version of the task, but the general effort level is not particularly high. Exposing participants to financial incentives motivates participants to raise the overall effort level substantially, but the revenue created for the principal decreases drastically. This counterproductive effect of performance pay is caused by the workers' decision to concentrate almost exclusively on the inefficient, easy version of the task when incentivized. Combining the fixed wage with a charismatic motivation speech, in contrast, increases both the overall effort level and the revenue for the principal. The positive effect on the effort level is smaller than the one of incentives, but the charismatic speech induces workers to focus on the difficult version of the task. A combination of incentives and a charismatic speech leads to similar outcomes as using incentives alone. These results establish novel insights: On one hand, we show that charisma can be an effective motivation tool even in situations where incentives fail. On the other hand, however, charisma does not shield participants from the corrupting effects of incentives when the two tools are combined in our setting.

How Narcissists Match and Play in Games
Evangelia Kyriazi

Why are narcissistic individuals overrepresented in influential positions such as CEOs and political leaders, despite the well-documented downsides of narcissism—poor teamwork, reduced cooperation, and lower organizational performance? This paper explores two main mechanisms that may explain this puzzle: partner selection and self-selection into competition. Through two controlled laboratory experiments, I investigate how individuals match, how they behave in strategic environments, and how they self-select in competitive environments depending on their own and others' levels of narcissism. The experiment uses three types of games—public goods (collaboration), ultimatum (fairness), and competition—to mimic core social and economic interactions. Participants complete the Narcissistic Personality Inventory (NPI), then are matched under three conditions: (1) random matching with no information, (2) random matching with partner personality info, and (3) directed matching via the Gale-Shapley algorithm, which lets individuals express preferences.
Findings reveal a consistent aversion to narcissism across all games. Even highly narcissistic participants prefer to be paired with less narcissistic individuals. This indicates that narcissism is perceived negatively and is avoided in these settings. In the public goods game, both one's own and the partner's narcissism level significantly affect behavior: narcissistic participants contribute less, and they also contribute less when matched with narcissistic partners. In the ultimatum game, narcissistic proposers make smaller offers, reinforcing the connection between narcissism and reduced fairness.
When people are allowed to self-select in competition, an interesting divergence emerges. I find that there are no narcissism differences in performance ex ante. Among the participants who choose to enter a competition, the probability of succeeding is higher for people high in narcissism because they self-select better. This difference is explained by the fact that low skilled and high in narcissism individuals choose not to enter the competition since this could be an ego-threatening situation. I do not find this difference in the luck task, where the ego threat is no longer active, and high and low skilled narcissistic participants do not differ in their selection.
These results suggest that narcissists are not necessarily more skilled but are more strategic in choosing when and where to compete. Combined with the widespread preference for non-narcissistic collaborators, this implies that narcissists rise to the top not because they are preferred, but because they self-select better into high-stakes environments.

Breaking Barriers: The Impact of Co-Leadership on the Gender Gap in Leadership Participation
Zahra Murad

Despite growing efforts to promote gender equality in the workplace, women remain significantly underrepresented in leadership positions across private and public sectors. For example, in 2021, only 5.6% of CEOs in Russell-3000 companies were women, and women's representation drops sharply along the corporate pipeline. Interestingly, women fare better on corporate boards, where leadership is collective—suggesting that shared responsibility may play a role in narrowing gender gaps in leadership participation.
While structural barriers contribute to this gap, emerging research shows that a gender difference in the willingness to lead is also critical. One overlooked psychological mechanism behind this reluctance is

responsibility aversion—the hesitation to assume leadership roles due to the anticipated psychological burden of being responsible for others' outcomes. We define responsibility aversion as distinct from general risk aversion or lack of confidence, and show it is especially salient for women in workplace leadership contexts.

In this paper, we investigate whether co-leadership—a leadership structure where two individuals share responsibility—can serve as an institutional solution to reduce gender disparities in leadership participation. We hypothesize that co-leadership can reduce psychological costs associated with leading, particularly among women, thereby increasing their willingness to lead. We present evidence from three controlled online experiments involving 2,700 participants and observational data from over 8,000 U.S. corporate boards. Our experiments show that women are significantly less likely than men to volunteer for solo leadership roles, with an average gap of 14 percentage points. When leadership is shared with a co-leader, however, this gender gap disappears. Women's willingness to lead increases, driven by a reduction in the psychological burden of responsibility. These findings are robust across tasks involving financial decision-making, redistribution, pro-social behavior, and performance-based representation. We also examine alternative explanations—such as risk preferences, confidence, and leadership motivation—but find no evidence that these account for the change in leadership willingness. Moreover, our observational analysis of U.S. corporate boards shows that women are more likely to serve as co-chairs than as solo chairs, consistent with our experimental findings. This research contributes to both the gender and organizational behavior literatures by identifying responsibility aversion as a key psychological barrier to women's leadership and by introducing co-leadership as a viable institutional solution. Unlike affirmative action or training programs that often ask women to adapt to traditionally masculine leadership norms, co-leadership reshapes the structure of leadership itself to make it more inclusive. Our results have practical implications for organizations seeking to diversify leadership without relying on quotas or tokenism. While co-leadership is not a one-size-fits-all solution, it provides a promising and underutilized pathway for reducing gender gaps in leadership roles. Future research should explore its broader effects on organizational dynamics, leader performance, and team outcomes.

Friday, September 19

| 09:00 – 10:30 | **Session 3L (Seminar Room B18.003)** |

Order effects in eliciting preferences
Sebastian Goerg

Having an accurate account of preferences help governments design better policies for their citizens, organizations develop more efficient incentive schemes for their employees and adjust their product to better suit their clients' needs. The plethora of elicitation methods most commonly used can be broadly distinguished between methods that rely on people self-assessing and directly stating their preferences (qualitative) and methods that are indirectly inferring such preferences through choices in some task (quantitative). Alarmingly, the two approaches produce systematically different conclusions about preferences and, therefore, survey designers often include both quantitative and qualitative items. An important methodological question that is hitherto unaddressed is whether the order in which quantitative and qualitative items are encountered affects elicited preferences. We conduct three, pre-registered, studies with a total of 3,000 participants, where we elicit preferences about risk, time-discounting and altruism in variations of two conditions: 'Quantitative First' and 'Qualitative First'. We find significant and systematic order effects. Eliciting preferences through qualitative items first boosts inferred patience and altruism while using quantitative items first increases the cross-method correlation for risk and time preferences. We explore how monetary incentivization and introducing financial context modulates these results and discuss the implications of our findings in the context of nudging interventions as well as our understanding of the nature of preferences.

Who Likes It More? Using Response Times To Reveal Group Preferences in Surveys
Carlos Alós-Ferrer

Surveys remain crucial tools for measuring societal preferences, but their reliability is limited by noise and bias in respondent data. We introduce a novel non-parametric method that leverages response times to reveal group preferences and rank preference strength across different populations. We validate the approach and apply it to key socio-economic questions using large representative surveys. The method complements traditional survey analysis techniques, providing clear indicators of when standard analyses may be inadequate and when response time data can yield additional insights. Importantly, our method also quantifies response biases, allowing researchers to adjust for systematic distortions in survey data.

Identifying Latent Intentions via Inverse Reinforcement Learning in Repeated Linear Public Good Games
Carina Hausladen

Robust results from public good games continue to defy theory. Uncertainty about the distribution of social preferences can explain first round contributions, but not the variance of contribution patterns in repeated play. Using a large-scale dataset (50,390 observations from 2,938 participants) we address this gap with two methodological contributions. First we propose a clustering approach (dynamic time warping) that reflects the nature of the data: it is two-dimensional, relating choices to experiences; and it allows changes to occur at idiosyncratic points of time. Second we refrain from constraining dynamic reactions to predictions derived from static social preferences. Instead, we treat reward function design as an estimation problem, using Inverse Reinforcement Learning to model behavioral patterns as discrete switches between latent intentions. This approach reveals that apparently noisy behavior in social dilemmas can be systematically explained as fluctuations between distinct latent intentions. Our framework

successfully accounts for behavioral patterns previously categorized as noise, providing a new paradigm for understanding dynamic decision-making in social dilemma games.

Session 3R (Seminar Room B18.005)

Norms in Collective Decision Making

Susann Adloff

We study the role of social norms for decision making in group choice contexts. Throughout live domains, many crucial decisions are made within groups, such that understanding group decisions becomes key to understand many societal outcomes. Social norms are a central driver of behaviour in individual choice contexts. Given the marked difference in the degree of pivotality of the individual decision maker in individual compared to group choice contexts, the role of social norms in group choce settings reamins unclear however. On one hand, a drop in pivotality can be associated with diffusion of responsibilty (lower norm following). On the other, it reduces the perceived cost of conformity (increasing norm following). We compare choices made in a classic dictator game setting, i.e. 1:1, with choices made in a modified version where the dictator's distribution proposition is matched with the propositions of 10 other randomly chosen dictators and the median proposition is implemented, i.e. 1(+10):1. Between these two contexts we compare the role of social norms by tracing social norm perceptions, the intrinsic importance of norm following and the effectiveness of extrinsic enforcement. In two further treatment conditions we aim to disentangle the role of diffusion of responsibility and low costs of conformity. Doing so, this study provides central evidence to better understand determinants of decision-making in group contexts and about the degree to which results on the role of social norms in individual choice contexts can be transferred to group decisions.

Cooperation and Communication in a Risky Social Dilemma: Experimental Evidence from the Field

E. Lance Howe

In small-scale personal exchange economies, communication is used to coordinate harvesting, hunting, and the pooling of resources to cover idiosyncratic shocks. We present results from a series of field experiments conducted in Western Alaska with indigenous Yup,Äôik and Cup,Äôik people who maintain strong social norms about the harvesting and sharing of subsistence foods (e.g. salmon, moose, and marine mammals). After identifying local norms of sharing, cooperation, and punishment through ethnographic fieldwork, we conducted a series of field experiments to test the effects of communication and reputation on cooperation in a risky social dilemma. We find that communication significantly increases cooperation in two domains, the traditional group investment domain and in the group self-help domain. Transcript data, in English and Yup'ik and Cup,Äôik languages, indicates that communication is used to address non-cooperative behavior in both sharing and investment domains. The effects of reputations are counter to expectations; the sharing of resources to pool risk is actually lowest in reputation treatments with and without communication. Consistent with our earlier findings from field experiments in Kamchatka Russia, subjects strongly condition sharing on the perceived worthiness of the needy player, with and without communication.

A good neighbor -- a found treasure: on the voluntary provision of public goods in overlapping neighborhoods

Juliane Koch

Public goods are often local in nature—ranging from air quality and green space to neighborhood safety—and their benefits depend on spatial proximity. Despite this, most public goods experiments abstract from space. A vast experimental literature has examined motives behind voluntary contributions to public goods (e.g., Andreoni, 1990; Zelmer, 2003), often using standard linear public goods games with fixed groupings and no spatial structure. These studies highlight behavioral motivations such as reciprocity (e.g., Falk et al., 2013) and inequality aversion (e.g., Cherry et al., 2005), but overlook how spatial positioning can affect contribution   decisions.

We design and implement a novel experimental framework to examine the voluntary provision of public goods in a spatially explicit context with overlapping neighborhoods, inspired by real-world debates around mixed vs. segregated urban environments. Our design is also informed by recent theoretical work on public goods provision in networks (e.g., Bramoullé & Kranton, 2007; Allouch, 2015) and complements experimental studies on multi-level or local-global public goods settings (e.g., Fellner & Lünser, 2014; Lange   et   al.,   2022).

Participants are located in a circular network and benefit from public good investments made at their own location and at those of their two immediate neighbors. This structure allows for a richer strategy space than canonical public goods games: individuals not only choose how much to contribute but also where to invest, determining who benefits. Our treatments vary both spatial neighborhood structure (closed vs. overlapping) and endowment heterogeneity (homogeneous, alternating rich-poor, and clustered rich-poor). The spatial distribution of inequality thus becomes endogenous to players' strategic and normative considerations.

In line with theoretical predictions, participants primarily invest in their own neighborhoods. Yet, we find robust evidence that both reciprocity and inequality concerns shape not only how much players contribute, but where they allocate contributions. Players are sensitive to the past behavior of their neighbors and adjust their investment locations accordingly—rewarding generous behavior and, conversely, shifting contributions away from underperforming neighbors. Moreover, under endowment heterogeneity, players direct more investments toward locations that benefit the poor, particularly when doing so benefits themselves   as   well.

Surprisingly, neither overlapping neighborhood structures nor endowment inequality significantly affect average payoff levels across treatments. However, the distribution of payoffs is meaningfully shaped by spatial configuration. In clustered settings, where rich and poor are spatially segregated, redistribution is largely driven by the rich players situated at the interface between clusters. These "rich border players" disproportionately contribute to their poorer neighbors, reducing inequality between clusters—but inadvertently creating inequality within their own. In contrast, mixed (alternating) settings exhibit less internal   stratification.

Our findings have both theoretical and policy relevance. They demonstrate that overlapping neighborhood structures allow for a more nuanced expression of social preferences, including conditional cooperation

and distributive concerns. Furthermore, the results suggest that promoting spatial integration—through mixed-income neighborhoods—can enhance equity outcomes without undermining efficiency. This experiment provides a lower-bound estimate of behavioral richness in real-world settings and opens avenues for future research on endogenous neighborhood formation, larger networks, and targeted redistribution mechanisms in spatial public goods.

11:00 – 12:30   **Session 4L (Seminar Room B18.003)**

Strategic Incompetence
Sorravich Kingsuwankul

We use a laboratory experiment to investigate the strategic use of incompetence where individuals communicate beliefs about their own abilities when an undesirable task has to be assigned to and performed by someone in a group. We vary the gender stereotype associated with the task (neutral, female or male) to examine the impact of stereotypes on the gender gap in strategic incompetence. If individuals believe that women (men) are more able at certain tasks, we can expect men (women) to claim more incompetence as those claims are perceived as more credible. In our experiment, participants first perform a trivia task after which we elicit their private belief about their ability. Participants then form a group of four to play a group investment game. We use a modified one-shot public good game with a cheap-talk communication where only one member can invest. The payoff structure is such that while it's better for the group that a high-type (i.e., top-2 performers in trivia of the group) invests, there is no incentive to do so as investing is costly. To decide who should invest, each player sends a cheap-talk message about the likelihood they believe themselves to be a high-type, then selects who should invest through majority voting. In this strategic setting, participants can lie to others by claiming to be a low-type. Indeed, we find that individuals strategically claim to be incompetent to avoid being assigned an undesirable task. When the task is stereotypical, the gender gap in strategic incompetence emerges but only with experience. Surprisingly, individuals distort private beliefs to be a low-type when the stereotype of the task is gender incongruent. Our work contributes to the literature on self-stereotyping by showing that individuals use stereotypes to justify claims of incompetence,Äìclaims that are partly motivated by the distortion of private beliefs in the direction of those stereotypes.

Gender Differences in Preferences: Means, variability, and malleability
Christian Thoeni

Despite extensive research on gender differences in economic preferences, the magnitude and relevance of these differences remain contested. Some argue that gender-based preference differences are small and of little predictive value, while others see them as important determinants of differential choices of men and women and potential explanations for gender inequalities in social and economic outcomes. We conduct a large-scale meta-analysis of experimental economics studies to establish how significant gender differences in behaviors relating to economic preferences are. To this end, we draw on a unique data set that combines the information of n=176 experimental studies with decisions from more than 130,000 subjects to analyze gender differences across five core preference domains: time, risk, competition, honesty, and social preferences. While mean differences vary across domains,Äîappearing robust for some preferences but negligible for others,Äîwe find evidence of greater male variability for a wide range of preferences. Moreover, men and women differ systematically in their responsiveness to treatment manipulations: men react more to economic treatment effects, while women respond more strongly to intra- and inter-personal treatment effects. We demonstrate the economic relevance of gender differences in preferences by showing that effect sizes of means differences in risk and competitive preferences are comparable to the size of the gender wage gap.

Reaction to Discrimination: Experimental Evidence
Vojtech Bartos

We study how Black Americans respond to observed harm against others in an online experiment (N = 4,477). We develop a Bystander Dilemma Game in which a White actor harms a passive individual whose racial identity (Black, Hispanic, or White) is randomized, and a Black bystander chooses to help, ignore, or harm. We benchmark bystander preferences using a Dictator Game with the same choice set but no perpetrator. We find strong ingroup favoritism and evidence of polarization: observing harm reduces average transfers but increases both very kind and harmful responses. Priming past discrimination strengthens ingroup support and reduces harm. The results highlight how observed discrimination can activate both solidarity and moral distancing within an increasingly multiracial society

---

Session 4R (Seminar Room B18.005)

Pluralistic Ignorance and Sustainable Mobility
Francesco Fallucchi

We examine the impact of goal setting and pluralistic ignorance on sustainable mobility choices through an experiment involving students. The objective is threefold: (1) investigate the influence of personal goal-setting on sustainable mobility; (2) determine whether informing participants about the percentage of people interested in improving their sustainable mobility habits increases such behaviour; and (3) explore the potential impact that increased use of sustainable mobility may have on psycho-physical measures of well-being. Preliminary results show that both personal goal-setting and information about other use of sustainable transportation means increase the share of sustainable mobility.

Label Uncertainty and Socially Responsible Market Behavior: An Experiment
Felipe Araujo

Socially-responsible (SR) labels (e.g., fair trade, carbon-neutral, etc.) have become ubiquitous in product markets. Corporations use SR labels to advertise their socially responsible behavior and attract SR-concerned consumers. A major challenge with such labeling schemes is the uncertainty regarding the minimum requirements (or standards) necessary to obtain each label. For example, with uncertainty around label standards, consumers might suspect that firms are taking actions to look socially responsible, instead of truly acting socially responsible, and adjust their behavior accordingly. We study the effect of label uncertainty using product market experiments. In a posted-offer market, sellers are randomly

assigned a product that either (a) imposes a large negative externality on a third party, or (b) a small negative externality, or (c) no negative externality. We also randomly assign a label standard in each round and sellers must decide whether to label their product (for a cost) as long as the minimum requirement is met. Specifically, with a very high standard only products with no externality can be labeled; with a moderate standard, products with either no negative externality or with a small negative externality can be labeled; and with a low standard, every product can obtain a label. In our uncertainty treatment, buyers observe if a product is labeled or not, but do not know the label standard for that round, whereas in our certainty treatment buyers observe both the label status and the label standard. We find that (a) buyers are willing to pay a higher price for labeled goods under both certainty and uncertainty; (b) buyers are willing to pay the highest difference for a labeled product when the label is perfectly informative of a low externality, (c) sellers with lower-quality products explore the uncertainty to obtain the (low standard) label and charge a higher price; and (d) the total amount of negative externality is lowest when there is no uncertainty about the label standards.

Beyond the Stars: Exploring the Welfare Effects of Ratings in Differentiated Markets
Luis Santos-Pinto

This paper examines the effectiveness of rating systems in guiding consumer choices in vertically and horizontally differentiated markets. Using a theoretical framework and an experimental market, we analyze how ratings affect consumer welfare. In vertically differentiated markets, where consumers agree on quality rankings, ratings improve welfare by directing consumers toward superior options. In contrast, in horizontally differentiated markets, where preferences vary across individuals, standard rating systems fail to enhance welfare. To address this, we test alternative mechanisms—filtered ratings and algorithmic recommendations—that improve consumer-product matching and increase average payoffs. Our findings underscore the role of rating design in its effectiveness and highlight how market structure and information aggregation shape consumer outcomes.

13:30 – 15:00   **Session 5L (Seminar Room B18.003)**

Unmasking the enemies: A theory of denunciations
Yu Tinghua

Denunciations are prevalent in authoritarian regimes. Citizens turn against each other to report suspicious behavior to the police state. But citizens may also have incentives to spread false information about their peers. In this context, can denunciations ever be informative? And, if so, what factors impede or facilitate the informativeness of denunciations? We design a formal model of denunciations in a large society. We show that denunciations are informative despite the certainty that some denunciations are false. Future works will highlight the complementarities between using informants and relying on denunciations for the secret police. We will also study how the regime can encourage denunciations and what it gains and potentially loses from incentivizing people to inform on one another.

Time Flies: Cooperation in Repeated Finite Time-Horizon Games
Petros Sekeris

In infinitely repeated games, the Folk Theorem guarantees that any feasible and individually rational payoff can be sustained as a Nash equilibrium. In contrast, finitely repeated games do not generally share this property. When the stage game has a unique Nash equilibrium, this equilibrium outcome is uniquely predicted at every stage through backward induction. This paper introduces a novel mechanism under which cooperation can arise even in finitely repeated games by embedding time misperceptions into the players' strategic reasoning.
We consider a finite-horizon repeated game with two players, where time is discrete and finite, $t = 1, ..., T$. Players start in a default action profile and may revise their strategies over time. The key departure from classical repeated games lies in how players perceive time: they may mistakenly believe the current period is earlier or later than it actually is. Formally, in each period t, players believe the game is in period $\tilde{t}$, where $\tilde{t}$ is a random variable supported on a neighborhood around t, capturing the idea that when "time flies" players make timing mistakes.
This misperception only persists as long as players adhere to the default action profile. If any player deviates, the illusion of temporal uncertainty is shattered and perfect time awareness resumes. Thus, cooperation becomes less risky when players are unsure about the current stage, as defection might come too early or too late relative to the intended timing.
We construct a simple two-player Prisoner's Dilemma example to illustrate the mechanism. Under standard assumptions, the unique Nash equilibrium of the stage game is mutual defection. However, if a Pareto-superior outcome exists, players may choose actions that implement this outcome as the equilibrium outcome throughout the finite game when time is perceived imprecisely. Players may be disciplined at equilibrium by anticipating that a deviation might prematurely or belatedly trigger punishments. We show that there exists a wide range of payoffs and belief structures under which cooperation strictly dominates unilateral deviation at every stage.
The theoretical contribution lies in showing that endogenous cooperation can emerge not from discounting or reputational concerns, but from the strategic consequences of temporal misperception. Our model departs from traditional imperfect monitoring approaches: here, players observe actions perfectly but are uncertain about when those actions take place. This simple shift in perspective allows us to recover cooperative outcomes in a finite game without assuming irrationality or repeated interaction beyond a known horizon.
In the general case, we characterize the equilibrium set under subgame perfection and identify the conditions under which Pareto-efficient strategies can be supported despite finite repetition. This work thus opens new avenues for understanding cooperation where agents operate under bounded temporal awareness, a scenario plausible in fast-paced or low-information environments.

Equilibrium Transitions in Sampling Dynamics: An Experimental Investigation
Yu Yaoyao

Understanding whether and how agents can move from inferior equilibria to superior ones is critical for improving social welfare. Evolutionary game theory provides a framework for studying transitions among

multiple strict Nash equilibria of games played by a population. The sampling best response dynamics (sBRD) (Oyama et al., 2015) predicts that more inexact information (i.e., a random sample of opponents' play) facilitates efficiency-improving equilibrium transitions. The equilibrium transitions in the sBRD are driven by sampling errors: agents observe random samples of the population play, which can be unrepresentative. Thus, a suboptimal action to the whole population could be a best response to a sample. Since sampling errors decrease with the size of a sample, the sBRD predicts that a small sampling size can facilitate equilibrium transitions. We test these predictions using a lab experiment.

In the experiment, a group of 14 subjects played a coordination game in one of three treatments differing in signals of opponents play: full information, a medium sample size of 7, and a small sample size of 2. We observe more efficiency-improving transitions with more exact information. We identify two channels for equilibrium transitions by combining experiments, structural estimations and counterfactual analysis. First, in all three treatments, some subjects exhibit strategic teaching behaviour i.e., the teachers intentionally deviate from myopic best responses to current signals towards an action of a more efficient equilibrium for higher long-run payoffs. Second, the counterfactual analysis shows that with no or a small number of teachers, the inexact information leads to more efficiency-improving transitions than full information.

However, the effect is smaller than theoretical predictions, because subjects use fictitious play i.e., they best respond to a history of observed signals rather than current signals only, especially under smaller sampling sizes.

Overall, the inexact information promotes equilibrium transition given a small number of teachers, while revealing full information works better when sufficiently many teachers are present.

Session 5R (Seminar Room B18.005)

Negative Emission Technologies and Climate Cooperation
Michela Boldrini

Negative Emissions Technologies (NETs) — a range of methods to remove carbon dioxide from the atmosphere — are a crucial innovation in meeting temperature targets set by international climate agreements. However, mechanisms that undo the adverse consequences of short-sighted actions (such as NETs) can fuel substitution effects and crowd out virtuous behaviors (e.g., mitigation efforts). For this reason, the impact of NETs on environmental preservation is an open question among scientists and policymakers. We model this problem through a novel restorable common-pool resource game and use a laboratory experiment to exogenously manipulate the key features of NETs and assess their consequences. We show that crowding out only emerges when NETs are surely available and cheap. The availability of NETs does not allow experimental communities to either conserve the common resource for longer or accrue higher earnings, and makes the earnings distribution more unequal.

Risk and Complexity of Climate Consequences Reduce Sustainable Behavior
Sebastian Olschewski

Human behavior is central to both causes and solutions to climate change. The consequences of one's actions, however, are often complex and uncertain, posing a potential barrier to sustainable behavior. While risk and complexity are well studied in financial decisions, little is known about the cognitive processes underlying sustainable behavior under such conditions. Here, we present three preregistered experiments examining how risk and complexity influence incentive-compatible sustainable decisions with real-world personal rewards and carbon emissions. Risk was implemented through two-outcome lotteries that either affected personal reward or climate consequences, while complexity was implemented through terms with additions and subtractions replacing single outcome numbers. We found that sustainable behavior, that is choosing the option with fewer carbon emissions, decreased when climate consequences became risky or complex compared to when personal rewards were risky or complex (effect of risk: $t(134) = 5.91$, $p < .001$; effect of complexity: $t(131) = 6.92$, $p < .001$). Complexity, but not risk, further reduced sustainable choices compared to decisions under certainty, $t(131) = 3.39$, $p < .001$. Computational modelling showed that risk/complexity aversion alone cannot explain this sustainability gap. Instead, participants systematically gave less weight to the climate attribute during decision-making when it was risky or complex. We discuss resource-rational adaptation and motivated cognition as explanations behind these effects. Our findings underscore the need for institutions to provide accessible environmental information to mitigate the disproportional effects of risk and complexity in information processing.

The acceptability of carbon taxation and regulation under inequality: An experiment
Assia Abdelfattah

Environmental measures are essential to reduce pollution, slow global warming and ensure the sustainable development of societies. Carbon taxes designed according to the Pigouvian tax framework constitute natural policy instruments for addressing environmental protection. (Timilsina, 2022). However, when framed as a uniform tax on the consumption of carbon emitting goods, such taxes can be perceived as regressive and unfair (Ewald, Sterner and Sterner, 2022; Klenert and Mattauch, 2016; Osman, Schwartz and Wodak, 2021;), can potentially crowd out intrinsic consideration for environment (Bazin, Ballet and Touahri, 2004) and can be perceived as inefficient leading to low social support (Bostrom et al., 2012; de Groot and Schuitema, 2012). Although the question of the progressivity of the tax system is essential in unequal societies, it is rarely studied in the case of taxes on consumption besides including redistribution. Additionally, a current debate is around the use of regulatory instruments such as efficiency standards or taxation to promote environmental protection (Levinson, 2019). Which type of instrument is most efficient? Are they equally acceptable?

To answer these questions, we run an incentivized stylized laboratory experiment. In total, 330 students participated in three between-subjects treatments depending on the policy instruments implemented. We test a uniform tax, a progressive tax and a consumption standard on the consumption of the polluting-good. In each treatment, participants either receive a high or low endowment when randomly assigned to heterogeneous groups. They decide on the quantities they want to consume from a polluting and a non-polluting good. The two types of goods have the same utility but differ in their costs and the negative externality emitted on other group members when consumed. The total consumption of the two goods is limited. The marginally decreasing utility functions are based on Cherry et al. (2012). In each treatment,

after 10 rounds of decision, the policy instrument is introduced on the consumption of the polluting good for 10 additional rounds. After the policy experimentation, participants have to vote on the policy instrument implementation at the beginning of each round, for 10 rounds. Parameters are such that in equilibrium, the three policy instruments lead to the socially optimal externality level. Preliminary results show that although participants gain from the implementation of the policy instrument, not all participants vote in favor of its implementation: on average less than 60% of the participants vote in favor of the flat or progressive tax while 74% vote in favor of the implementation of standard. Interestingly, low- and high-endowed participants vote equally in favor of the implementation of the standard, they have different behavior regarding the implementation of the different taxes. Our study provides results on the impact of policy instruments on the reduction of negative externalities and its acceptability among consumers depending on their income. Previous studies consider only one good and homogenous consumers. Our experiment contributes to the literature considering the acceptability of policy instruments under unequal contexts.

15:30 – 17:00   **Session 6L (Seminar Room B18.003)**

Behavioral Foundations of Financial Resilience in the Digital Economy: Evidence from Switzerland and Singapore
Sandra Andraszewicz

Understanding the behavioral and institutional drivers of financial resilience is critical in an era where individuals increasingly interact with complex financial systems mediated by technology. This study explores how individual attributes—such as personality traits, psychological resilience, and cultural orientation—interact with formal (e.g., financial systems, digital infrastructures) and informal (e.g., social networks, lifestyle norms) institutions to shape financial resilience in technologically advanced economies. We investigate these dynamics using large-scale, cross-national survey data from Switzerland and Singapore—two developed, digitally integrated, yet culturally and institutionally distinct countries. We collected data from the general populations in both countries (Switzerland N = 2,388; Singapore N = 2,068) and from a FinTech-savvy user base of a Swiss mobile payment app (TWINT, N = 1,347). These data allow us to investigate the links between psychological and financial resilience with FinTech adoption, and to examine how individual traits and institutional contexts jointly influence financial behavior. Financial resilience—the ability to absorb and recover from financial shocks—is not only a function of income, financial literacy and behaviour, or access to financial services, but is also predicted by psychological resilience, including mental health, stress recovery capacity, and optimism. Using random forest analysis and regression modeling, we find psychological resilience to be the a robust behavioral determinant of financial resilience.
Social resilience—defined as the strength and utility of social networks and embeddedness in the local context—also facilitates financial resilience. Individuals embedded in strong social networks report higher access to informal financial support and show greater capacity to adapt to financial adversity. Moreover, cultural orientation—specifically, the extent to which individuals feel emotionally connected to the country they live in—positively predicts financial resilience. This effect is consistent across both Switzerland and Singapore, suggesting that social cohesion and identification with the institutional environment foster cooperative behaviors that enhance economic stability.
FinTech users, across both countries, are on average more financially and psychologically resilient than the general population. Importantly, we find that adoption of FinTech tools is not solely driven by objective assessments of utility or ease of use, as proposed in classic Technology Acceptance Models. Instead, adoption is predicted by perceived lifestyle fit. This underscores the need for regulatory models that account for subjective risk perception and behavioral heterogeneity, rather than relying solely on normative rational-choice assumptions.
These findings reveal a complex interplay between behavioral traits and institutional context infacilitating financial wellbeing. By integrating psychological and social dimensions into economic models of financial behavior, we contribute to a more comprehensive understanding of human behaviour and financial resilience in increasingly digitalized financial systems. Our data is available as an Open Science "Resilience and Tech Database" which is a living project regularlly updated with datasets in other countries and on specific FinTech users.

The Invisible Infrastructure: Rethinking Care, Vulnerability, and the Economy
Valentina Rotondi

Care is the invisible infrastructure of our economies: essential to human development and social cohesion, yet chronically undervalued in economic models and policy frameworks. This paper develops a new conceptual and formal framework in which care—when sustained as a public responsibility—unlocks the productive potential of human vulnerability. We model how institutional framings of care as either private or public shape not only well-being and productivity, but also individuals' willingness to redistribute time and resources to sustain caregiving systems. Building on this model, we conduct a pre-registered survey experiment (N = 450) testing whether brief narrative interventions can shift beliefs about the public nature and productivity of care, and thereby increase support for redistributive care policies. Participants are randomly assigned to one of three framing conditions (neutral, private care, public care), and their beliefs and redistributive preferences are measured both before and after treatment. Results show that exposure to a public framing of care increases both the perceived value of care and willingness to contribute to caregiving infrastructure—through taxation and time. These findings offer a new behavioral mechanism through which narratives influence economic preferences, and provide policy-relevant insights into how societies can build institutional support for care by reshaping what people believe it is for.

The Price of Identity: Overoptimism and Congruence Concern
Michael Kurschilgen

We examine how identity influences economic decision-making, using field experiments on sports betting to measure belief distortions and identity-driven preferences. We find that people overestimate the likelihood of identity-aligned outcomes by 10%–18%, and allocate 20% more of their betting budget to teams for which they have an affinity than to neutral teams. Using a structural model of portfolio allocation, we show that overoptimism accounts for 30%–44% of this investment gap, while the remaining 56%–70% stems from an aversion to betting against one's favored team, even when such bets offer higher expected

returns. Our estimates suggest that this aversion is equivalent to discounting gains from identity-incongruent outcomes by 17%–27%. We also provide evidence for the "identity-threat response" theory: when individuals perceive their identity as under threat -- such as after their team's poor performance -- they strengthen their commitment, reinforcing identity-driven betting. Our findings raise policy concerns, as identity-driven biases may exacerbate financial harm not only in the rapidly expanding sports betting market but also in broader consumer and financial decision-making contexts where identity affects choices.

Session 6R (Seminar Room B18.005)

Fighting Wasteful Spending through Inter-Unit Transfers in Organizations
Fidel Petros

In many public institutions, year-end budget rules encourage inefficient spending: unspent funds are forfeited, leading to wasteful last-minute purchases. While rollover policies can mitigate this inefficiency, they are often politically or legally restricted. This project proposes and experimentally tests an alternative: allowing units to transfer unspent budget to others. Using a controlled laboratory setting that mimics multi-year public budgeting under uncertainty, I compare three regimes: Expiring, Rollover, and Transfer. I then explore under what institutional conditions transfer becomes effective, focusing on reciprocity and symmetry, or the lack of any of them. The experimental design isolates the policy rule from the implementation mechanism, enabling clean identification of behavioral responses. By introducing inter-unit transfers as a potential second-best solution, the project contributes to the literature on budgeting and institutional design based on behavioral mechanisms.
Preliminary results from pilot data will be available at the time of the symposium (Pilot will be run on July 15).

Subjective performance evaluation in tournaments with asymmetric contestants: Experimental Evidence
Francisco Martinez

We theoretically and experimentally study the consequences of principals suffering from correlation neglect, focusing on subjective performance evaluations. We derive a Lazear-Rosen tournament model involving a tournament administrator (the principal) and two agents competing for a monetary prize. Each agent exerts effort (unobserved by the principal) with mean-zero added noise. One agent has an exogenous advantage (in our setup a project-specific case), receiving an additional asymmetric shock that positively influences their work output. Principals have full information about this induced asymmetry, and their goal is to maximize agents' efforts by selecting an optimal allocation winning rule based on the total work output of each of the agents. Theoretically, the solution for setting optimal incentives is straightforward: Principals can create a fair tournament by giving a handicap to the exogenously advantaged player equal to the magnitude of the exogenous shock times its relative contribution to the total work output. This ensures that, in expectation, the hardest-working agent wins the tournament, which maximizes principals' payoff. However, principals who suffer from correlation neglect may neglect the correlation between the agents' output and the exogenous shock. Consequently, against their own best interests, they may fail to design the optimal promotion rule because they are unable to infer the relevant underlying parameter from observed correlated signals. We further derive agents' best responses and show that when principals suffer from correlation neglect, the two agents' optimal effort provision decreases due to the discouragement effect.
Our experimental data reveals that principals suffer from correlation neglect: they are unable to (fully) disentangle effort from the exogenous asymmetric shocks. Principals learn over time and the inefficient allocation rule improves over rounds. We also document that the principals' biased promotion rules have tangible consequences in terms of inequality: Advantaged agents systematically earn more due to their higher winning chances. Therefore, principals' correlation neglect creates an inequality in payoffs. To conclude, despite the distorted incentives, agents do not adjust their effort provision downwards, although they believe they face unfair tournaments.

Information valuation and workers' selection into formal or informal transactions: a lab-in-the-field experiment with street vendors in Bogotá
César Mantilla

A challenge in developing economies is the co-existence of formal and informal labor markets. The usual discussion is whether this is a problem of exclusion, with rigidities and market failures preventing workers from being absorbed by the formal market; or a problem of voluntary exit from this formal market because the interactions with the state are unsatisfactory. This dissatisfaction can arise from higher costs of formality (e.g., bureaucracy, tax payments) compared to the benefits. However, this dissatisfaction does not necessarily come from experience but also from the negative priors about these interactions. I explore how this negative perception may result in a barrier to formal transactions. I measure it in a simple manner, embedding it in the payment receipt. The participants-303 street vendors in Downtown Bogotá, representing informal workers-are asked whether they are willing to forgo part of their study earnings to avoid signing a receipt and writing their ID number.
Giving up money to hide basic personal information, as the national ID number, can be thought as a behavioral entry barrier to formal transactions and deters interaction with other formal institutions and more elaborate formal procedures. Since this behavior can be rooted in the lack of trust, we use one of the archetypical paradigms, a trust game. In this game, the first-mover's behavior is interpreted as trust and the second-mover's behavior is interpreted as trustworthiness. Since I am interested in trust, and how trust may depend on the members of society with whom street vendors interact, all the 303 participants played as first-movers. They interacted with three types of anonymous and previously collected second-movers from whom we made their identity salient: police members patrolling their area, public officers from the local Mayor Office, and other street vendors.
I find that in about thirty percent of transactions, participants preferred not to sign a receipt, even if this was costly in terms of the received payment. In some choices, to avoid signing the receipt, and in addition to the payment deduction, participants must provide either their phone number (to pay throughout their mobile wallet) or allow the enumerators to take a picture of the moment when they get paid (as an alternative proof of payment). Yet, the proportion of participants who prefer not to sign the receipt remained in about

one-third. This opens important questions about which type of information the street vendors may consider more valuable, sensitive, or concealable. Regarding the trust game, we find trust rates of around 50% when interacting with the two out-group identities (police and public officers), and trust rates of around 70% when interacting with a member of their in-group. Surprisingly, those who trusted more in other street vendors were less likely to sign the receipts, while trusting the out-group is not correlated with signing the receipt. These results remark on the need to better understand this reluctance to engage in simple procedures associated with formal transactions, and how informal networks and their reliance on trust can be self-reinforcing.

17:15 – 18:15 **Keynote Lecture (Aula)**

Query Theory: A Process Account of Constructed Judgments and Preferences
Elke Weber

Query Theory (QT) describes the process of generating and aggregating internal evidence about the merits of different judgment or choice options for the purpose of arriving at a judgment or choice. Its assumptions about selective attention and goal-directed memory retrieval provide a unified explanation of a broad set of behavioral anomalies. QT assumes that boundedly-rational decision makers issue sequential queries to their episodic and semantic memory that request arguments in support of a specific choice option, one option at a time, but switching between options. The returned evidence is aggregated, and the option with greater support at the time of decision is selected. Critically, the request for evidence supporting the first-considered option temporarily inhibits arguments for all other options that are response competitors, making evidence supporting later-queried options harder to access when subsequent queries turn to them. Normatively irrelevant features of the way a choice is presented can be shown to influence initial attention and hence query order and choice, including incumbency or status-quo (which option is the behavioral or recommended default), surface appeal (an attractive label or image), or mechanics of the choice situation such as presentation order or reading order. The causal role of query order is established by prompting respondents in a QT study explicitly to reverse the order of queries from their "natural" order (i.e., the order aligned with the framing condition) to the reverse or "unnatural" order, which greatly attenuated the effect of frame on judgment or choice. A meta-analysis of 27 QT studies confirms all QT assumptions and shows that an average effect size of .34 for the effect of decision frame on query order. QT processes have also been shown to mediate the influence of social norms on attitudes and behaviors, where norm information directs initial attention on norm-congruent arguments and thus influences policy support.

| 09:00 – 10:30 | Session 7L (Seminar Room B18.003) |

Whodunit: Individualistic Remedies Undermine Support for Systemic Reforms
David Hagmann

Policymakers and public interest groups often treat individual and structural reforms as complementary routes to societal change. We show, however, that calling attention to individualistic remedies can systematically suppress public demand for the systemic policies that complex problems typically require. Across five preregistered experiments (N = 6,690) spanning climate change, obesity, retirement savings, and gender inequity, we demonstrate a robust "responsibilization" effect: exposure to solutions framed around personal behavior decreases (i) the perceived responsibility of government and firms, (ii) the rated importance of structural interventions, and (iii) revealed preference support for organizations lobbying for systemic change.

Building on cognitive research that people attend to only one construal of a problem at a time, we predict that spotlighting personal actions crowds out consideration of situational causes and policy levers. This mechanism extends attribution bias theory by showing how frames that highlight the actor (the "i frame") divert attention from the environment (the "s frame"), thereby altering downstream preferences over real policy choices.

Each study randomly assigned online U.S. participants to read brief prompts (news headlines or policy vignettes) emphasizing either (a) individual behavior, (b) systemic reforms, or (c) neutral facts (in Study 4). Participants then: (1) Proposed their own solutions (open ended, bonus incentivized); (2) Rated the relative responsibility of individuals versus government on 7 point scales; (3) Indicated which approach they deemed most important using a continuous slider, and (4) In Study 3, allocated a $200 real donation between a financial literacy charity (individual intervention) and a lobbying group for pension reform (systemic intervention).

Our key findings show a shift in attribution, policy importance, and behavioral choice. Across domains and studies, the Individual frame reduced the likelihood of proposing systemic solutions for the problem and shifted the relative attribution of the responsibility for solving the problem from governments to individuals. The same frame lowered ratings of systemic reforms. Individuals primed with individual solutions were less likely to direct the $200 donation to the systemic reform NGO. We show this effect replicates even if participants do not first write an open-ended solution (Study 2). Study 4 includes a neutral condition and shows that the effect is driven by a shift away from systemic solutions in the individual conditions, with participants naturally proposing systemic solutions. Finally, the effect replicates even when the individual and systemic primes present information that a particular solution was ineffective.

These results reveal an unintended cost of widely touted behavioral change interventions and actions proposed by public interest advocates: even modest, well intentioned messages about personal responsibility can undermine public appetite for high leverage structural change. Practically, communicators seeking transformative policy should avoid exclusive focus on individual fixes and instead integrate them within narratives that keep systemic levers salient. Our work advances behavioral public policy scholarship by quantifying the cognitive trade off between i and s frames and offers a diagnostic tool for anticipating when good faith individualistic messaging may backfire against long term systemic progress. It also suggests that behavioral interventions ("nudges") focusing on individual actions may come at an unexpected cost to systemic reforms.

Activating Change: The Role of Information and Beliefs in Online Activism
Ahana Basistha

Online activism has become an important form of civic engagement, yet little is known about what drives participation or how different forms of activism respond to the same appeal. Using data from a pre-registered online experiment with nearly 2,000 participants in India, in collaboration with a local NGO fighting healthcare corruption, we examine how information about the social problem and beliefs about others‚Äô support influence engagement in three forms of activism: signing a petition, donating money, and watching an advocacy video. Participants were randomly assigned to receive information about the cause or about others' support, both, or neither, and were cross-randomized into action conditions. By examining multiple forms of online activism within a unified experimental framework and context, we can compare how individuals respond to distinct calls to action. We find that petition signing increases in response to both information about the issue and belief-correction about others' support. In contrast, donations and video viewing are largely unaffected, and donations may even decline. Additionally, allowing participants to choose among types of activism reduces engagement. These results suggest that information and belief correction can increase participation in low-cost actions, and limiting activism options can boost engagement.

The Effect of Reflection on Loan Decisions
Sara Khayouti

Microcredit was once seen as a powerful tool for reducing poverty. Yet a growing body of evidence shows that microcredit often does not lead to large or sustained improvements for the poor (Banerjee et al., 2015; Meager, 2019). This suggests that access to loans is only part of the challenge. Many individuals in low-income settings face behavioral frictions that lead to limited attention, imperfect recall of business operations, or a lack of structured planning (Mani et al., 2013; Kremer et al., 2019). We argue that these behavioral frictions may prevent them from utilizing loans efficiently, and we empirically investigate how a simple reflection intervention, designed to improve structured thinking, can improve small business owners' loan and business decisions.

We study our research question with entrepreneurs in Dar es Salaam, Tanzania, who apply for loans from a large microfinance company. We ask whether prompting borrowers to reflect systematically on their business operations at the time of application affects their borrowing behavior, repayment rates, business performance, and subsequent loan applications. We evaluate the intervention through a randomized controlled trial in the field with 129 loan applicants in the pilot of our study. While prior work has shown that reflection can improve learning and judgment in various settings (Di

Stefano et al., 2015; Hauser et al., 2014), our study is, to our knowledge, the first to rigorously test whether induced reflection influences borrowing behavior in a real-world microfinance context. Related work by Augenblick et al. (R&R, QJE) document that segmenting memory can improve consumption smoothing decisions in seasonal poverty contexts. Our study differs as it examines production-side decisions (borrowing for business purposes), uses a low-cost and scalable phone-based intervention, and includes a more detailed reflection module.

Preliminary results suggest that the reflection treatment did not significantly affect loan sizes but led to modest improvements in early repayment behavior. Treated clients were somewhat more likely to revise their requested loan amount, though these differences were not statistically significant. Repayment performance, however, was consistently higher in the treatment group across the first few weeks of the repayment cycle.

Our intervention offers a simple, low-cost, and scalable tool that has the potential to improve the lives of poor entrepreneurs in low-income settings. By encouraging borrowers to reflect more carefully on their business operations and loan planning, the intervention can reduce default rates and support more informed borrowing decisions. This benefits not only the clients, who may avoid over-indebtedness and stressful repayment, but also loan officers, by reducing the time they spend on loan recovery as well as lowering opportunities for corruption. Once proven effective, the reflection tool can be easily rolled out across other branches and institutions, making it a practical policy solution with wide applicability in microfinance.

Session 7R (Seminar Room B18.005)

The behavioral foundations of international anti-bribery laws: Results from an international lab-type experiment

Theodore Alysandratos

This study employs experimental methods across nine nations (N = 2506) to scrutinize the behavioral foundations of international anti-bribery laws. It assesses how individuals' conduct is influenced by their expectations of enforcement, specifically gauging the probability that a 'Monitor' will detect, report, and penalize their transgressions.

The international community has invested considerable resources in building international anti-corruption legal architecture, to which many governments have signed up. This includes the wide-ranging UN Convention Against Corruption, which has 190 state parties, and the OECD Anti-Bribery Convention, where 46 states have agreed to outlaw the payment of bribes to foreign officials. However, historically, countries have varied considerably in the implementation of their commitments and in the resources they have devoted to enforcement. A particular feature of those laws is their extra-territorial applicability, meaning that they potentially impact a large swathe of international business. While there is scant evidence about the impact of such laws on behavior, a growing body of research suggests that individuals engage in corruption where they see it as a widespread social norm and/or because they assume that others in their environment will behave corruptly.

We report the findings from an experimental study involving bribery-like exchanges, opportunities to punish those choosing to initiate such exchanges across international borders, and opportunities to evaluate the various decisions that could be made in this context. The experiments were designed to provide answers to the following six questions. First, in the interests of establishing a link between the experiment and what is already known about corruption globally, we ask:

RQ1. Does the social appropriateness of bribe offering, bribe taking, and exposing bribers to the threat of punishment vary across countries and in accordance with prior expectations?

Then, we move on to our primary questions:

RQ2. Does the presence of international monitors who can initiate processes leading to the punishment of international bribers reduce international bribery?

RQ3. Does the efficacy of such international monitors depend on their location?

And, finally, we delve into mechanism:

RQ4. What role do beliefs about the likelihood of monitors using their powers play in determining the monitors' efficacy?

RQ5. How do potential international bribers' beliefs vary depending on where those who are monitoring them are located? and

RQ6. Are investors' beliefs about the likelihood of monitors using their powers, conditional on the monitors' locations, correct?

The findings reveal a substantial and significant decrease in the incidence of bribery when a Monitor is present. This decline results from the perception that the Monitor will expose corruption, despite participants harboring inaccurate beliefs about the likelihood of Monitors imposing penalties. Our results suggest that the extra-territorial enforcement of anti-corruption laws, such as the Foreign Corrupt Practices Act and the new Foreign Extortion Prevention Act, holds promise in combating corruption in international business transactions. However, our findings indicate that effective deterrence depends on authorities widely publicizing their enforcement activities. The desired impact of new laws cannot materialize if those targeted lack accurate beliefs about the authorities' inclination to enforce them.

Standing in Prisoners' Shoes: A Randomized Trial on How Incarceration Shapes Criminal Justice Views

Aljosha Henkel

Over the past few decades, many regions across the world have experienced a notable shift toward more punitive criminal justice policies. This trend is most evident in the widespread rise of mass incarceration, which has resulted in nearly 11 million confined individuals worldwide (Fair and Walmsley, 2021). While public opinion is considered a key factor in shaping criminal justice policies (Baumer and Martin, 2013; Brace and Boyea, 2008; Enns, 2016; Jennings et al., 2017; Nicholson-Crotty et al., 2009), most individuals possess very limited knowledge about the everyday realities of life behind bars (Bryant and Morris, 1998; Doble, 1995; Farkas, 1993; Hough and Roberts, 2005; Roberts and Hough, 2005). Prisons, by their very nature, remain largely inaccessible for citizens, leaving public perceptions confined to media portrayals and fiction (Cecil and Leitner, 2009; Wilson and O,ÄôSullivan, 2004). This raises the questions of whether public attitudes towards incarceration are rooted in an accurate understanding of prison life, and whether

society‚Äôs stance on criminal justice policies would remain unchanged if people experienced incarceration firsthand?

In this paper, we shed light on these questions by taking advantage of a unique randomized controlled trial that provided regular citizens with an opportunity to gain firsthand incarceration experience. The Department of Justice and Home Affairs from the Canton of Zurich in Switzerland organized a four-day test run for a newly built prison to test its operational and security procedures under simulated real-life conditions. Volunteering participants could experience up to 48 hours of incarceration, following standard procedures closely mirroring a real inmate‚Äôs journey from intake to release, including strip searches, confiscation of all belongings, and a strictly regulated daily routine. Given that the number of interested volunteers exceeded the prison‚Äôs capacity, we randomly divided applicants into two groups. The prison authorities were free to select any applicant from the treatment group and schedule a slot in the test run, while we instructed them not to invite applicants from the control group. This design allows us to estimate the intention-to-treat effect of having the opportunity of gaining firsthand incarceration experience. We conducted surveys with applicants before and after the test run to gauge their punitive attitudes and beliefs about wellbeing in prison, using both behavioral and questionnaire measures.

Our empirical results first reveal that, in contrast to the control group, subjects who received the opportunity to take part in the test run became substantially less likely to support punitive policies. Moreover, they became more likely to donate money to organizations advocating for a more moderate, rather than a tougher, criminal justice system. Second, although we observe that the public has very inaccurate beliefs and substantially overestimates the wellbeing of prisoners, the opportunity to participate in the test run did not alleviate their misperceptions and the treatment group did not significantly change their beliefs compared to the control group.

Overall, this randomized controlled trial offers novel causal insights into the impact of firsthand incarceration experience and underscores the crucial role of personal experience in shaping support for criminal justice reform.

Institutional rules and unfair rule enforcement

Simon Columbus

Laws prescribe formal sanctions to deter rule violations. Early economic analyses of the law assumed that the expected cost of such formal sanctions must outweigh the benefits of breaking the law to be deterrent. However, laws also have an 'expressive' function, signalling or establishing social norms about what is considered appropriate and what is not. Consequently, even laws prescribing weak or infrequent formal sanctions may reduce noncompliance by signalling social norms. Conversely, sanctioning institutions may fail to signal, or even undermine, social norms if they are perceived as biased. We specifically focus on the effects of institutional bias in how rule violations are monitored. Such cases are common: for example, Black taxpayers in the US are substantially more likely to be audited than non-Black taxpayers. However, despite the evident significance of this form of bias, we know of few systematic studies of its effects on rule compliance.

We examine the causal effects of institutional rules and biased rule enforcement on compliance and cooperation in heterogeneous groups. Our experimental design is based on the standard public goods game. We introduce heterogeneity by establishing minimal group identities, that is, we randomly assign players to either 'red' or 'blue' subgroups. We then vary the presence of a contribution rule and its enforcement, and observe the effects on costly contributions to the public good, rule compliance, and norms. First, we introduce a non-binding minimum contribution rule. Comparing treatments with and without this rule allows us to estimate the expressive function of non-binding rules. Second, we introduce non-deterrent sanctions for rule violations. In the fair audit treatment, all players are audited with the same probability and fined if found in violation of the rule. This allows us to distinguish the effect of sanctions from the expressive effect of the rule itself. Finally, in the biased audit treatment, 'red' players are three times more likely to be audited than 'blue' players, though all players face the same fine if found in violation of the rule. Comparing treatments with fair and biased audits then allows us to estimate the effect of biases in rule enforcement on rule compliance and contributions to the public good.

We conduct a large, incentivised online experiment (N = 1,254). The presence of a minimum contribution rule is associated with a substantial increase in contributions to the public good. Moreover, the presence of a rule has a strong and positive effect on personal and social norms – consistent with the expressive function of rules – and sustains contributions over time. Introducing sanctions for rule violations further increases compliance, but does not carry over into an increase in average contribution levels. Instead, the positive effect of rule enforcement on the compliance of free-riders is almost perfectly offset by a reduction in contributions above the level stipulated by the rule. Finally, biased audits initially decrease compliance relative to fair audits, but this difference diminishes over time. Thus, our results suggest that the expressive function of rules is remarkably robust to biases in how rules are enforced.