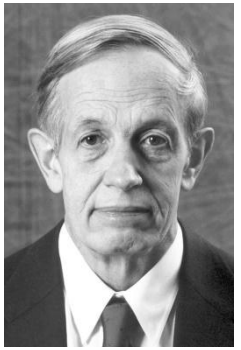
Preferenced Based Games & Nash Equilibrium Play

ESA Europe Meeting 2025

Hannes Rau, KIT Karlsruhe

*joint with F. Kauffeldt (Heilbronn University) & A. Sofianos (Durham University)

Motivation



- **Nash Equilibrium**: Widely used concept in Economics and many other (social) sciences
- In several applications, Nash-equilibrium fails to accurately predict agents' behaviour
- Sometimes players even choose strictly dominated strategies (e.g. cooperating in a Prisoner's Dilemma/PG game)
- Violation of rationality? Rather unlikely.

→ If not, there must be some other relevant factors missing in the analysis…

Relevant Factors Missing

- In many applications, Nash prediction based on players' own material payoffs (observable & measurable)
- But in many cases agents do not only care about own payoffs, but also about *payoffs of others* and the *way of interaction* (*"social preferences"*)
- **As a consequence, persons' utilities of the outcomes of the games may differ substantially from own material payoffs** (whereas the latter usually is used for the equilibrium prediction)

→ Agents in fact may face a very different strategic situation than the originally described game (+missing common knowledge)

Example: Prisoner's Dilemma

- Material payoffs normalized to values 1-4
- **Case 1, two selfish players:** $(4, 1) \succ (3, 3) \succ (2, 2) \succ (1, 4)$
- Preference Game corresponds to the **standard Prisoner's Dilemma** with **one pure Nash equilibrium**

	D	C
D	2, 2	4, 1
C	1, 4	3, 3

- **Case 2, two cond. cooperators:** $(3, 3) \succ (4, 1) \succ (2, 2) \succ (1, 4)$
- Preference Game corresponds to a **Coordination Game (Stag Hunt)** with **two pure Nash equilibria**

	D	C
D	2, 2	4, 1
C	1, 4	3, 3

Our Research Questions

- **How does the category of a game* change, when accounting for the (social) preferences of the players?**

→ Comprehensive analysis for all 2x2 games

- **Do rates of equilibrium play improve when using the transformed games for the prediction instead of the original ones?**

→ Further study planned to answer this RQ

* game structure based on own material payoffs

Previous/Related Work

- **Rates of Nash equilbirum play substantially increase when incorporating social preferences into the prediction**
  (Rau, 2024, Working Paper)

- **Rates of equilibrium play significantly increase when agents have common knowledge about the actual game being played**
  (Brunner, Kauffeldt & Rau, 2021, EER)

(Experimental) Framework

- Focus on simultaneous one-shot 2x2 games
- **Individual payoff values (x, y) are normalized to values 1-4*** for both players (no ties)

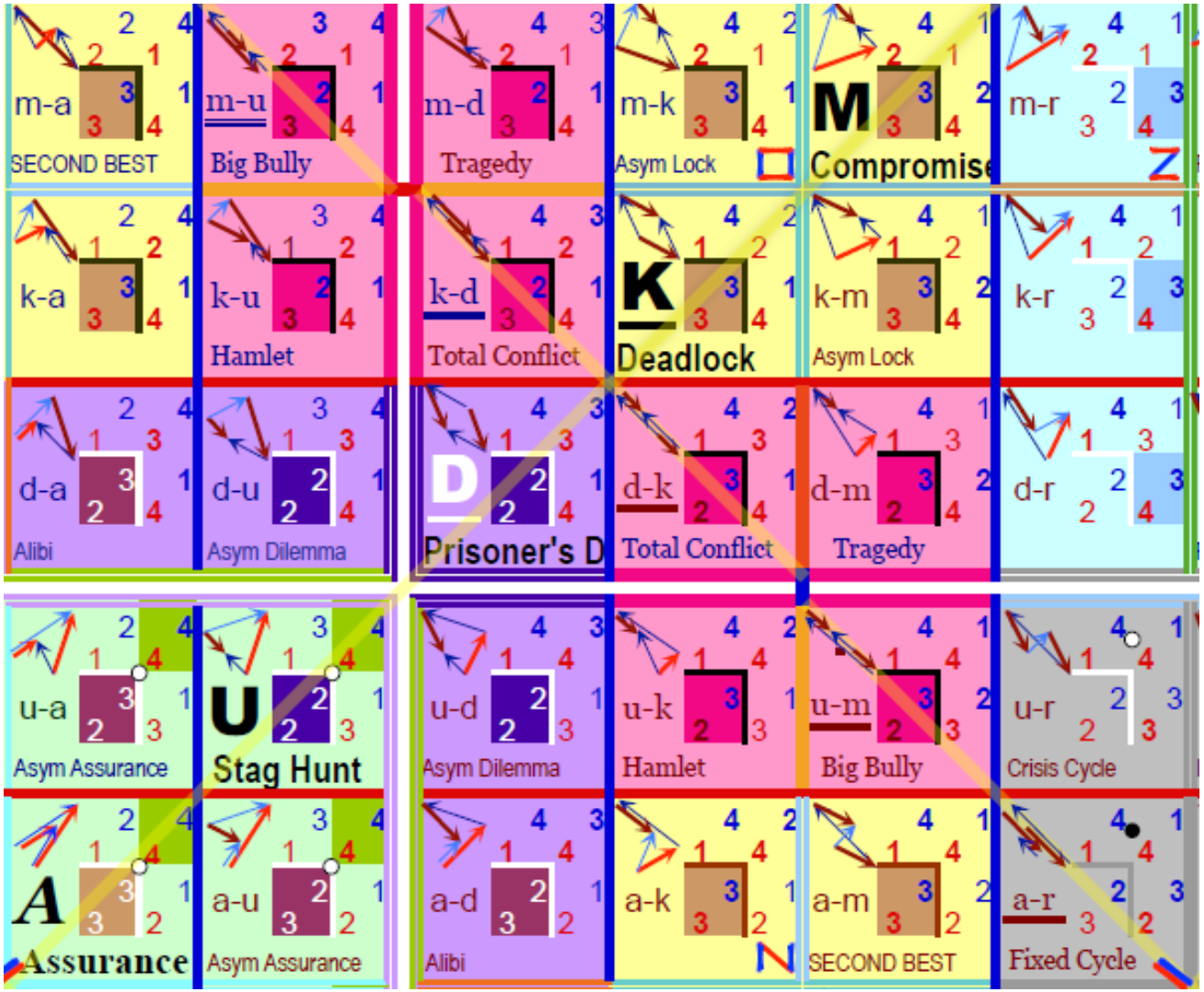
→ Yields 78 distinct strategic classes of 2x2 games

- **Elicitation of ordinal preference ranking over set of all sixteen possible payoff combinations (x, y)** (incentives for truthful reporting provided)
- Assumption (at first): **Consequentialism**

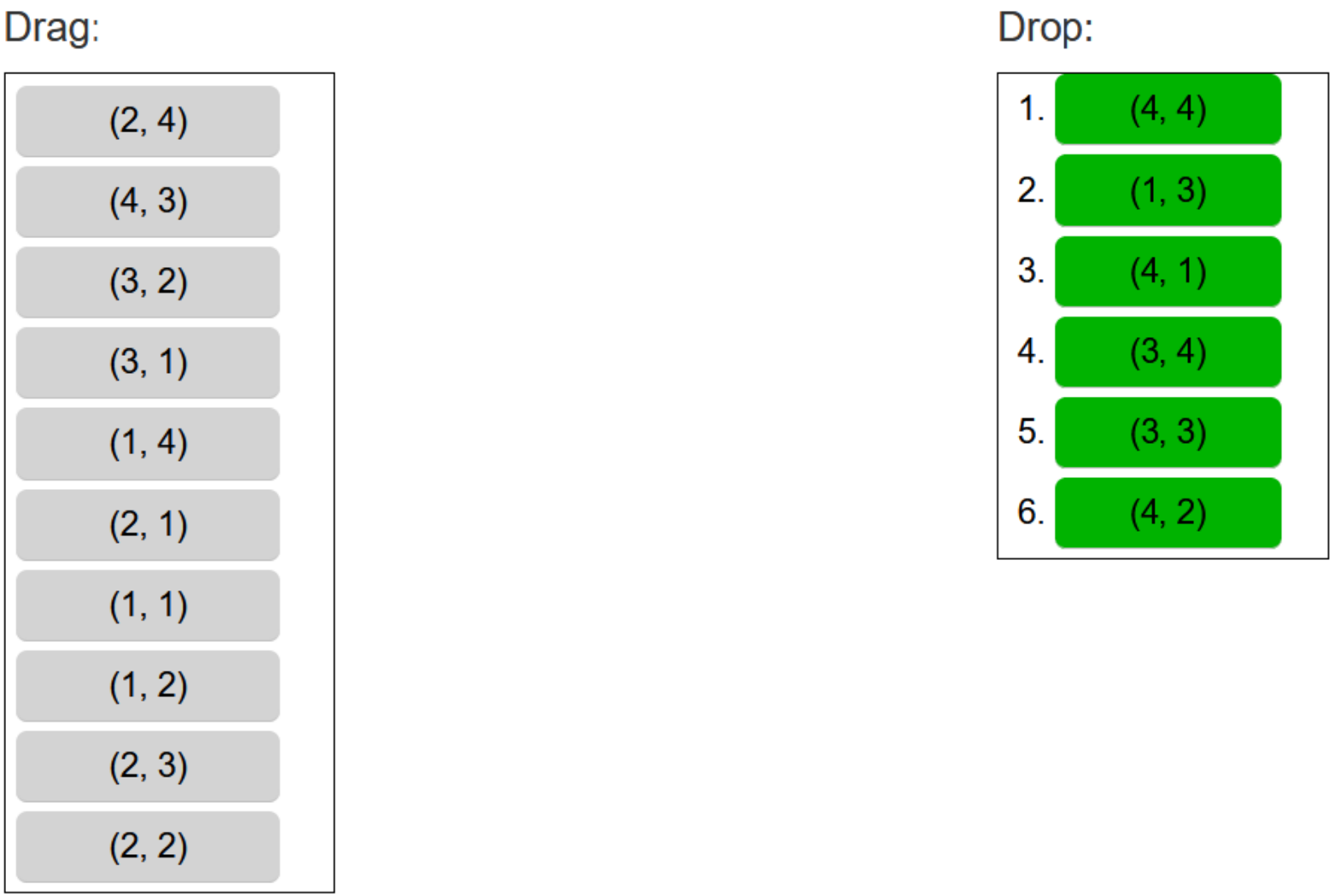
→ **Individual rankings can be used to identify (ordinal structure of) preference-based Games for any specific pairing of players**

*Robustness check with multiples of 1-4, e.g. {4, 8, 12, 16}

Games Classification (acc. to Bruns, 2012)



Screenshot: **Ranking of Payoff Combinations**



Method of Analysis

- Each original game contains 4 different payoff pairs (=outcomes of the game) from the set of 16 pairs
- The outcome ranked best among those 4 pairs receives the highest value for the respective player (again, values of 1-4 are used) →**mapping from monetary payoff vectors to utilities**

--> Same framework as before, but numbers in individual cells may be different for the players than before

--> May yield another game than the original one

--> Analysis is conducted for all 78 individual games and all potential (hypothetical) pairings of subjects (round robin matching)

--> For each initial game, this yields a distribution of transformed games which might possibly result from the original "monetary" game

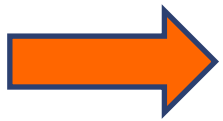
Method of Analysis: Example

Consider analysing the PD as original game and an agent with **cond. cooperator** type preferences: $(3, 3) \succ (4, 1) \succ (2, 2) \succ (1, 4)$

Outcome in original game	Assigned (ordinal) utility value in target game
(3, 3)	4
(4, 1)	3
(2, 2)	2
(1, 4)	1

Original Game (material payoffs)

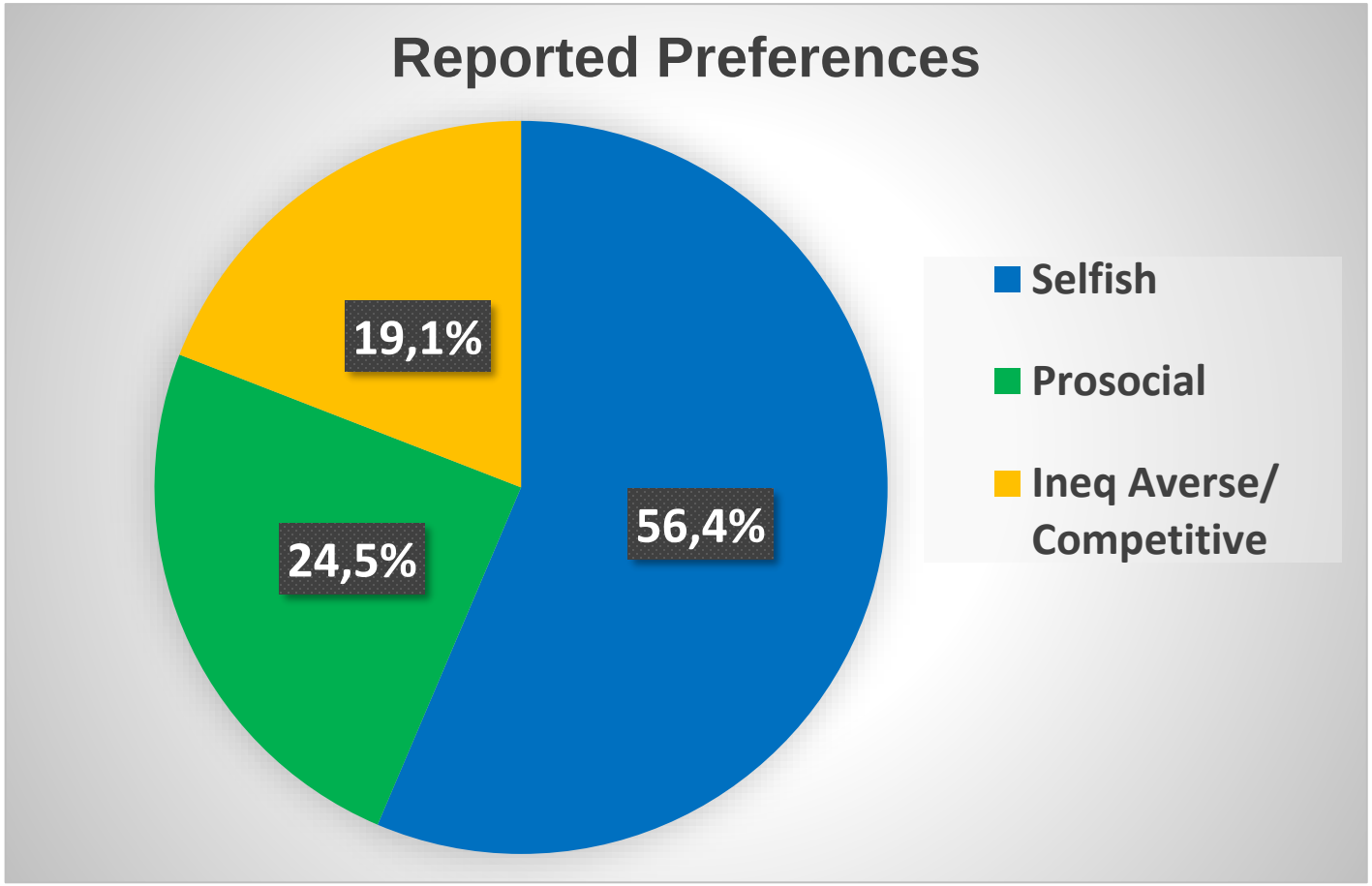
	D	C
D	2, 2	4, 1
C	1, 4	3, 3



Transformed Game (utility values)

	D	C
D	2, 2	3, 1
C	1, 3	4, 4

Results



n=212 subjects, 1 ranking per subject

Most common ranking: $(4, 4) \succ (4, 3) \succ (4, 2) \succ (4, 1) \succ (3, 4)$…

Results

- 78 distinct simultaneous 2x2 games
- Focus on ordinal structure and pure equilibria only
- Percentages how often original game (monetary game) differs from corresponding preference game: **3,7% - 53,8%**
- Games with the highest* number of changes (each 53,8%):
  - **Crisis Cycle**
  - **Deadlock**
  - **Prisoner's Dilemma**
  - **Second Best**

	A	B
A	1, 4	4, 1
B	2, 2	3, 3

	A	B
A	1, 4	2, 2
B	3, 3	4, 1

	A	B
A	1, 4	3, 3
B	2, 2	4, 1

	A	B
A	1, 4	4, 1
B	3, 3	2, 2

[*Because pair (3, 3) often is preferred to (4, 1)]

Results: Specific Games (Example)

- Original game: **Prisoner's Dilemma**

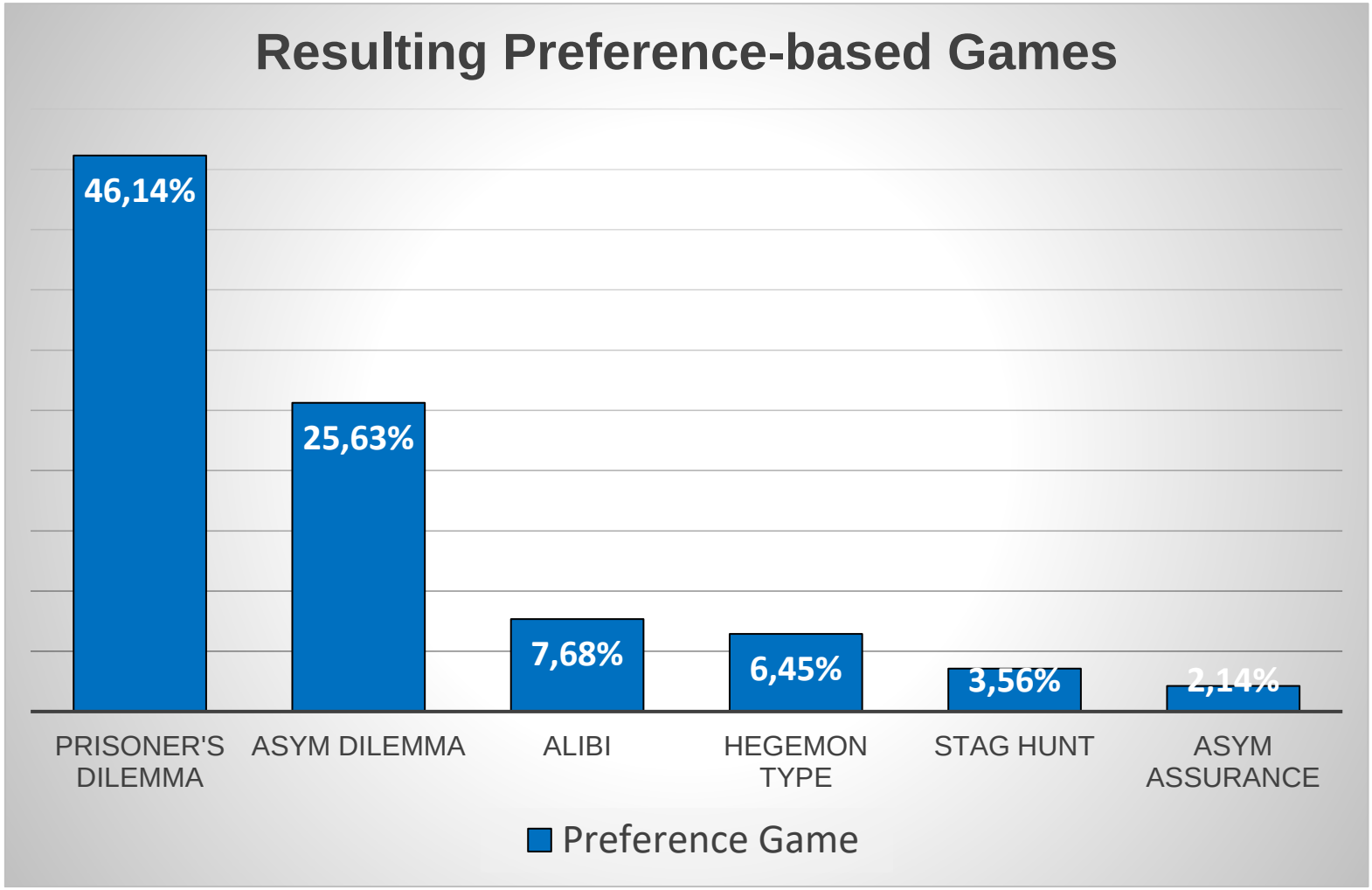
	A	B
A	1, 4	3, 3
B	2, 2	4, 1

- Observed rankings (from most to least often):
  - $(4, 1) \succ (3, 3) \succ (2, 2) \succ (1, 4)$ (selfish)
  - $(3, 3) \succ (4, 1) \succ (2, 2) \succ (1, 4)$ (prosocial/cooperative)
  - $(3, 3) \succ (2, 2) \succ (4, 1) \succ (1, 4)$ (inequality averse)
  - $(3, 3) \succ (4, 1) \succ (1, 4) \succ (2, 2)$ (max. total payoff)
  - (…)

Results: Prisoner's Dilemma

Possible pairings:	Resulting preference game:
selfish vs. selfish	Prisoner's Dilemma (d)
selfish vs. prosocial	Asym Dilemma (d-u)
selfish vs. inequality averse	Alibi (a-d)
selfish vs. max total	Hegemon Type (d-n)
prosocial vs. prosocial	Stag Hunt (u)
prosocial vs. inequality averse	Asym Assurance (a-u)
prosocial vs. max total	Anticipation (n-u)
inequality averse vs. ineq. av.	Assurance (a)
inequality averse vs. max total	Mutual (a-n)
max total vs. max total	No Conflict (n)

Results: Prisoner's Dilemma



n=44.944 (potential) pairings, round robin matching

Summary/Conclusion

- Many players exhibit social preferences
- Therefore, (Nash) equilibrium prediction based on own material payoffs often does not perform very well

→ **Our method allows for assessment which games are more likely to be affected by changes and what are the distributions of resulting Preference-based Games**

→ Useful for analysts/experimenters to known when one should take into account social preferences of the players

→ **Comprehensive analysis for all 2x2 games (ordinal structure)**

[For game play: Technically Bayesian framework/common knowledge needed → analysis soon gets very complicated]

Next Steps/Follow Up Study

Our current plan for the design:

- Two treatments with two stages (randomized order)
- Stage 1: Ranking of 16 payoff pair (as before)
- A) Stage 2: **Elicit rankings of outcomes in context of games**
  (8-10 selected games, some with high/average/low number of transformations from previous analysis)
- B) Stage 2: **Game play in selected games**
- Plus belief elicitation about opponent's rankings/strategies
- Robustness check: Variation of parameter values

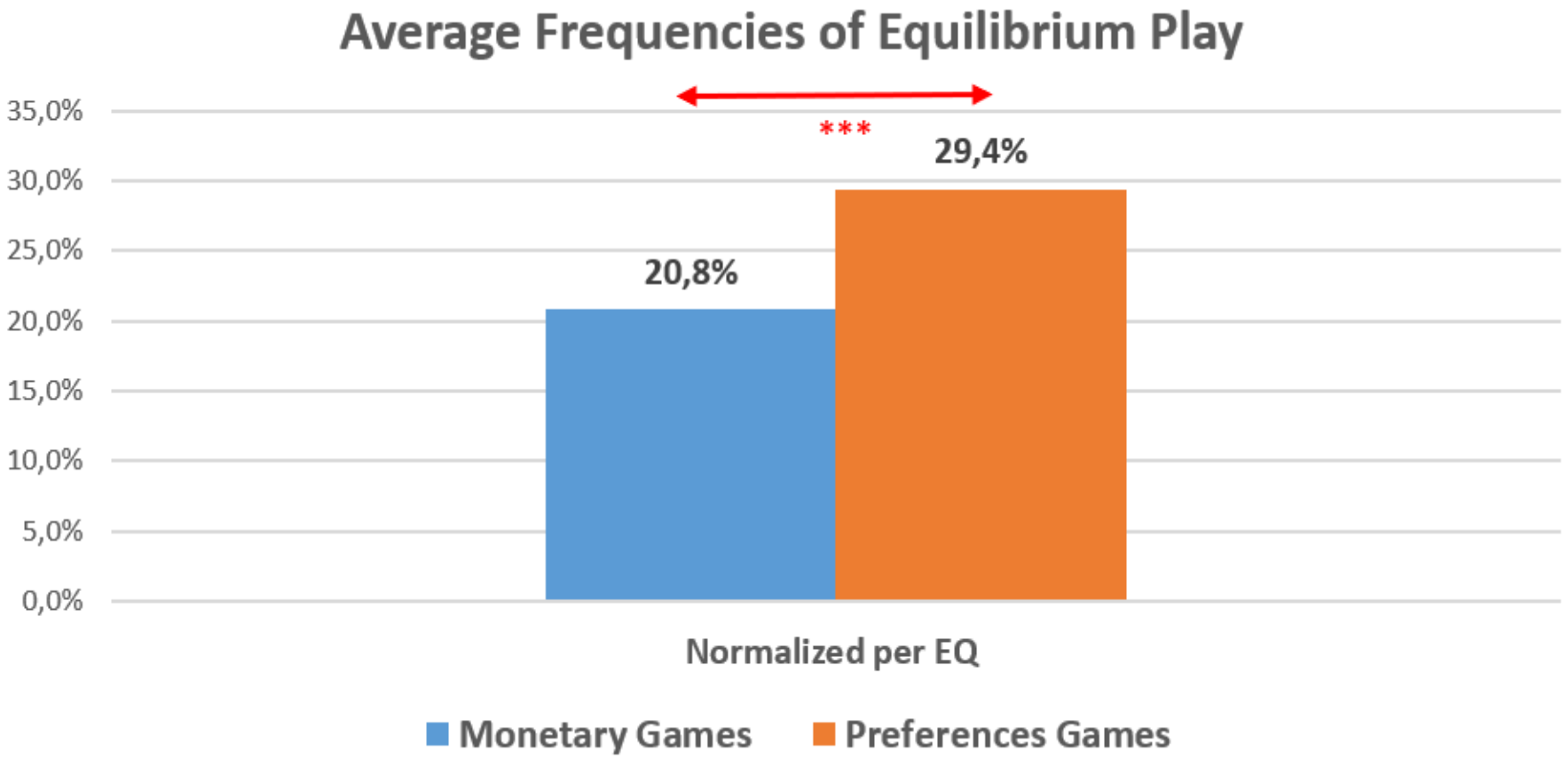
→ **Allows to test for consistency of ranking of pairs vs. outcomes of games (consequentialism)**

→ **Allows to test if subjects more often play a Nash equilibrium**


Thank
you

Karlsruhe Institute of Technology

Main Results (Previous Study)



n=188 subjects / 752 decisions; 53% of non-selfish preferences; mostly slightly prosocial; frequencies normalized by no. of equilibria existing)