Game theoretic approaches to social norms

Vincent Buskens, Wojtek Przepiorka, Werner Raub

Conference on Social Norms and Institutions
Centro Stefano Franscini, Ascona
May 10 – 15, 2015
Motivation

- Social norms are a central concept in social science, but are mostly taken for granted → *homo sociologicus*

- This approach bypasses the problem posed by Hobbes why there isn’t a war of all against all (Coleman 1964)

- Assuming *homo economicus*, we are interested in explaining how social norms emerge, are maintained and change.

- We argue that game theory provides a useful framework for theorizing about social norms.
Outline

- Part 1 (Wojtek): Notions of social norms; (tacit) emergence of social norms
- Part 2 (Werner): (Maintenance of) social norms of cooperation through repeated interactions, network formation and effects of networks
- Part 3 (Vincent): Social norms of cooperation through endogenous peer punishment
**Social norm**

*Definition*: Rule guiding social behavior, the deviation from (adherence to) which is negatively (positively) sanctioned.
Examples of social norms I
Examples of social norms II
Examples of social norms III

Please offer this seat to someone who needs it more than you do.
Some categorizations of norms

Social norms are directed at focal actions performed by target actors and benefit the norm beneficiaries.

- Proscriptive and prescriptive norms
- Disjoint and conjoint norms
- Moral, social, and legal norms
- Conventional and essential norms

But how do norms come into existence?
Emergence of norms

Necessary condition: **Externalities of actions** for which markets in rights of their control cannot be established **create a demand for a norm.**

Social dilemmas

**cooperators**  
**free-riders**

Definition: A situation of strategic interdependence in which the decisions of individually rational actors lead to an inferior outcome for all or some parties, than the decisions of “collectively rational” actors.

Who walks the dog?

Externalities
The volunteer’s dilemma (VOD)

Public good $\sum U_i$ for a group $n \geq 2$ is produced by one person $i$ choosing C at a cost $K_i$ where $U_i > K_i > 0 \ \forall \ i$.

Symmetric VOD: $U_i = U_j$ and $K_i = K_j \ \forall \ i \neq j$

Asymmetric VOD: $U_i \neq U_j$ and/or $K_i \neq K_j \ \exists \ i \neq j$
Social norms and repeated games

- Notion of social norms gains momentum if actors are repeatedly confronted with structurally similar social dilemmas.
- Repeated interactions allow for the emergence of (equilibrium) behaviour at which norms can be directed.
- Problem: There are many equilibria that can emerge in repeated social dilemmas.
Repeated games and the folk theorem

2-person sym. VOD

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>3, 3</td>
<td>3, 8</td>
</tr>
<tr>
<td>A2</td>
<td>3, 8</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

2-person asym. VOD

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>3, 5</td>
<td>3, 8</td>
</tr>
<tr>
<td>A2</td>
<td>8, 5</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

\[ \bar{u}_i = \sum_{t=0}^{\infty} \delta^t u_i(s(t)) \]

minmax point based on the mixed strategy equilibrium

\[ p_i^* = 1 - \frac{U_i}{K_i} \left( \prod_{j=1}^{n} \frac{K_j}{U_j} \right)^{\frac{1}{n-1}} \]

How can we do better than minmax without an external "choreographer?"
An exogenous correlating device could orchestrate agents’ equilibrium play and social norms could be conceived of as such correlating devices (Bowles & Gintis 2011, 89-92).

However, ‘social norms cannot be introduced as a deus ex machina,..., without violating the objective to provide a “bottom-up” theory of cooperation that does not presuppose preexisting institutional forms of cooperation.’ (Bowles & Gintis 2011, 90)

→ At present, the question how social norms emerge is also an empirical question.
Hypotheses

**H1**: In the symmetric VOD, actors will be more likely to take turns in cooperating than to coordinate on only one actor cooperating throughout.

**H2**: In the asymmetric VOD with one strongest actor, actors will be more likely to coordinate on only the strongest actor cooperating than on taking turns in cooperating.

(see Diekmann & Przepiorka 2015 for details)
Laboratory experiments
Experimental design

<table>
<thead>
<tr>
<th></th>
<th>Symmetric</th>
<th>Asymmetric 1</th>
<th>Asymmetric 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 3$</td>
<td>$U_i = 80c$</td>
<td>$U_i = 80c$</td>
<td>$U_i = 80c$</td>
</tr>
<tr>
<td></td>
<td>$K_i = 50c$</td>
<td>$K_{1,3} = 50c$</td>
<td>$K_{1,3} = 50c$</td>
</tr>
<tr>
<td></td>
<td>$K_2 = 30c$</td>
<td></td>
<td>$K_2 = 10c$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1 (N = 120)</th>
<th>Experiment 2 (N = 87)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partner matching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random matching</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Random assignment on conditions within session
- VOD in groups of 3 for 48 – 56 rounds (told 30 – 60)
- Partner m.: fixed groups and roles throughout
- Random m.: random groups and roles each round
In dieser Runde sind Sie Person 2.

Sie treffen Ihre Entscheidung, indem Sie mit der Maus in das mit "oben" oder "unten" beschriftete, blau unterlegte Feld klicken.

Entscheiden Sie sich für "oben", erhalten Sie:
- 70 Rp wenn Sie Person 2 sind.
- 30 Rp wenn Sie eine andere Personen sind.

Entscheiden Sie sich für "unten", hängt Ihr Gewinn von den Entscheidungen der anderen Personen ab:
- Entscheidet sich mindestens eine andere Person für "oben", erhalten Sie 80 Rp.
- Wenn sich alle Personen für "unten" entscheiden, erhalten alle 0 Rp.
1 Person hat sich für "oben" entschieden. Diese Person erhält 70 Rp.

2 Personen haben sich für "unten" entschieden. Diese Personen erhalten 80 Rp.

Sie haben sich für "unten" entschieden und erhalten 80 Rp.
Pattern: partner matching, sym.
Patterns: partner matching, asym. 1

Group members' decisions across sessions (2 = 'strong', x = 'cooperation')
## Results: summary

<table>
<thead>
<tr>
<th></th>
<th>partner matching</th>
<th>random matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>symmetric VOD</td>
<td>49.5 %</td>
<td>N/A</td>
</tr>
<tr>
<td>turn-taking</td>
<td></td>
<td>no pattern</td>
</tr>
<tr>
<td>asymmetric1 VOD</td>
<td>34.9 %</td>
<td>94.1 %</td>
</tr>
<tr>
<td>solitary volunteering</td>
<td></td>
<td>solitary volunteering</td>
</tr>
<tr>
<td>asymmetric2 VOD</td>
<td>61.7 %</td>
<td>95.0 %</td>
</tr>
<tr>
<td>solitary volunteering</td>
<td></td>
<td>solitary volunteering</td>
</tr>
</tbody>
</table>

- Possible explanation: inequality aversion
- Partner matching: the strong are exploited by the weak
- Random matching: costs of cooperation are equally shared
Are these social norms?

According to Lewis (1969), a convention is a behavioural regularity that emerges in a recurrent coordination game and establishes an equilibrium.

Lewis (1969: 97) acknowledges ‘conventions may be a species of norms: regularities to which we believe one ought to conform.’

In how far are conventions normative?
Normativity of conventions

Announced payoffs rounds 1 - 10

<table>
<thead>
<tr>
<th></th>
<th>Red</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>10, 10</td>
<td>0, 0</td>
</tr>
<tr>
<td>Blue</td>
<td>0, 0</td>
<td>10, 10</td>
</tr>
</tbody>
</table>

Change in payoffs round 10, but only player 1 is informed

<table>
<thead>
<tr>
<th></th>
<th>Red</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>200, 200</td>
<td>300, 0</td>
</tr>
<tr>
<td>Blue</td>
<td>300, 0</td>
<td>200, 200</td>
</tr>
</tbody>
</table>

Round | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Player 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>Player 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only 29% of players 1 deviated from the convention in round 10.
(see Guala 2012 for details)
Latent norms

‘[I]nteraction generates habits; perceived, they become reciprocal expectations; in addition to their purely predictive and anticipatory nature, sensitivity to them endows them with a constraining or even an obligatory character. [...] Thus do norms grow in unplanned fashion out of ongoing interaction.’ (Wrong 1994, 48-49)

‘The more often the members of a group perform a behaviour, B, the stronger is the empirical expectation that B is performed; the stronger this empirical expectation is, the stronger is resentment in case of non-regular behaviours; the stronger this resentment is, the more likely it is that the performance of B becomes a norm.’ (Opp 2004, 14)
Conclusions

- Game theory provides a framework for the analysis of situations in which externalities produced in social interactions create a demand for social norms.

- Social norms can emerge tacitly from behavioural regularities and latent norms in repeated encounters.

- (Repeated) game theory has some serious limitations (e.g. many equilibria, availability of information)

- Moreover, social norms can also emerge via direct communication, bargaining and social exchange. \( \rightarrow \) cooperative game theory

- A demand for a social norm is a necessary but not a sufficient condition for the emergence of a norm: (1) sanctioning mechanism, (2) norm proliferation.
References


Diekmann, Andreas, and Wojtek Przepiorka. 2015. ""Take One for the Team!" Individual heterogeneity and the emergence of latent norms in a volunteer’s dilemma." Working Paper, ETH Zurich.

Elster, Jon. 2007. *Explaining Social Behavior: More nuts and bolts for the social sciences*. Cambridge: Cambridge University Press. [Ch. 5, 8 & 22]


Norms of cooperation through repeated interactions, network formation and effects of networks
Overview: norms of cooperation in social dilemmas

- Focus: cooperation in social dilemmas
- “Demand” for norms of cooperation in social dilemmas: each actor is better off when everybody cooperates than when everyone defects, while there are individual incentives to defect
- “Realization” of norms: in repeated social dilemmas, following a norm of cooperation can be in the actors’ “enlightened” self-interest in the sense that following the norm is equilibrium behavior (self-enforcing norms)
- Norms of cooperation: social norms (i.e., deviations are sanctioned by actors’ peers); conjoint norms (i.e., target actors are also norm beneficiaries)
Overview: repeated interactions, networks, and norms of cooperation

- Repeated interactions can support norms of cooperation in social dilemmas
- Networks reinforce the effect of repeated interactions
  - Network effects
  - Games on networks
- Hence, actors have incentives to invest in networks
  - Strategic network formation
- Core feature: integrated model of network formation and effects of networks
  - We assume full strategic rationality, also with respect to network formation
- Focus on theory formation and formal modeling, including testable predictions
Brief detour on network effects and network formation
Network effects I

- Theory and empirical research show: social networks have effects for micro-level individual behavior as well as macro-level phenomena, e.g.:
  - Search behavior on the labor market; labor market outcomes (Granovetter)
  - Individual adoption and diffusion of innovations (Coleman)
  - Trust in exchange and behavior in social dilemmas (Coleman, this part of the workshop)
- Actors benefit from occupying certain individual positions in a network and from certain network structures
  - E.g., in some contexts (social dilemmas), dense networks are beneficial, while open structures are beneficial in other contexts (when access to and control of information is important)
Network effects II

- Tool for analysis of network effects: games on networks
- Networks and their characteristics are assumed to be exogenous
Network effects III

Reputation and Efficiency in Social Interactions: An Example of Network Effects

Werner Raub and Jeroen Weesie
Utrecht University

Reputations emerge if an actor's future partners are informed on his present behavior. Reputations depend on the “embeddedness” of interactions in structures or networks of social relations. They illustrate the effects of such embeddedness on the outcomes of interactions. This article presents simple game-theoretic models of reputation effects on efficiency (in the Pareto sense) in interactions. In a comparative perspective, the authors start with a baseline model of a social system in which reputation effects (of a specific kind) are excluded: actors do not receive information on their partners’ behavior in interactions with third parties. Such a system of “atomized interactions” is compared to a system with interactions that are “perfectly embedded”: actors are immediately informed on all interactions of their partners with third parties. Efficiency is more easily attained as a result of individually rational behavior in perfectly embedded systems. In a final step, the comparative perspective is broadened, and the extreme assumptions of either an atomized or a perfectly embedded social system replaced. Intermediate cases arise in the consideration of “imperfect embeddedness,” that is, a situation in which actors are informed only after some time lag on the behavior of their partners vis-à-vis third parties. It is shown that the conditions for efficiency become more restrictive as the information time lag lengthens.

In this paper we use a theoretical strategy that has been described by James Coleman (1988, p. S97) as an attempt “to import the economist’s social science...”
Network formation I

- Often, actors can affect their position in a network and the network structure, at least to some degree, by establishing, maintaining or severing relations with others
- Network effects & opportunities to affect one’s position in a network and the network structure
  - incentives to invest in networks
- Strategic network formation
- Networks and their characteristics are endogenous
Network formation II
Network formation III

Dynamics of Networks if Everyone Strives for Structural Holes¹

Vincent Buskens
Utrecht University

Arnout van de Rijt
State University of New York at Stony Brook

When entrepreneurs enter structural holes in networks, they can exploit the related benefits. Evidence for these benefits has steadily accumulated. The authors ask whether those who strive for such structural advantages can maintain them if others follow their example. Burt speculates that they cannot, but a formal demonstration of this speculation is lacking. Using a game theoretic model of network formation, the authors characterize the networks that emerge when everyone strives for structural holes. They find that the predominant stable networks distribute benefits evenly, confirming that no one is able to maintain a structural advantage in the long run.

INTRODUCTION

The view that social networks are a form of capital because they can facilitate economic activity is now generally accepted. The structure of an individual's social environment has been shown to matter in a number of ways. Job search through weak ties is more successful (Granovetter
Formation and effects of networks I

- Next step and new contribution: integrated model of strategic network formation and network effects
- Co-evolution of networks and behavior
- Context: social dilemmas
- Another new feature: we assume full strategic rationality, also with respect to network formation
Formation and effects of networks II

**Macro**
- Social conditions, including network characteristics

**Micro**
- Preferences, information

**Equilibrium behavior**
- Individual effects:
  - Behavior in social dilemma, including norm-following behavior
  - (Investments in) network formation

**Pareto-optimality; maintenance of a norm of cooperation; network formation**
What follows?

- Model assumptions
  - We consider a simple special case and sketch generalizations later on
- Analysis of the model: theorems
- Testable predictions
- Limitations of the model, generalizations and extensions, summary
Model assumptions
Trust Game as an example

- The model covers a class of social dilemmas
- We sketch the model using the Trust Game as an example of a social dilemma
- “Simple cases first”: a triad with *one* trustee and *two* trustors; focus on investments in and effects of a tie that allows for information exchange between the trustors
- Generalizations to be sketched afterwards
Trust Game

Trustor

No trust

Trustee

Trust

Abuse

Honor

\[ \begin{pmatrix} P_i \\ P \end{pmatrix} \quad \begin{pmatrix} S_i \\ T \end{pmatrix} \quad \begin{pmatrix} R_i \\ R \end{pmatrix} \]

\[ S_i < P_i < R_i \]

\[ P < R < T \]
Trust Game and norms of cooperation

- Actors have individual incentives to abuse trust and not to place trust
- Actors are better off when trust is placed and honored ("cooperation") compared to the no trust-situation
- Trust Game as a social dilemma
- Demand for a norm of cooperation
- What are conditions for realization of a norm of cooperation?
Repeated Trust Games I

- Consider a game $\Gamma$ with $n = 2$ trustors and one trustee
- In $\Gamma$, each trustor plays a standard Trust Game (TG) with the trustee in rounds $t = 1, 2,...$
- In each round $t$, the trustee plays first a TG with trustor 1, then with trustor 2
  - Note: we could assume that the sequence of trustors changes between rounds
Repeated Trust Games II

- Payoffs: $R_i > P_i > S_i$ for trustors and $T > R > P$ for trustee
  - Payoffs for trustors $i$ and $j$ may differ
  - We do not assume “symmetry” in the sense of $R_i = R$ and $P_i = P$
  - $T, R, P$ and $R_i, P_i, S_i$ are constant for all rounds $t$
- After each round $t$, the next round $t+1$ is played with probability $w$; $\Gamma$ ends with probability $1 - w$ ($0 < w < 1$)
- Hence, the trustee plays an indefinitely often repeated TG with each trustor
Revised Trust Games: example

- Consider two buyers as trustors who purchase goods from a seller as trustee under conditions of asymmetric information.
- Rounds of \( \Gamma \) are business periods (e.g., market days) in which each buyer decides whether or not to purchase a good from the seller and, if the buyers decide to buy, the seller decides on selling good quality or selling bad quality for the price of good quality.
- With probability \( 1 - w \), each period is the final one in the sense that the seller stops business due to some exogenous contingency beyond his control (for example, a new competitor enters the market who offers a superior product).
  - \( w \) represents the “shadow of the future” (Axelrod): with increasing \( w \), actors' long-term incentives increase.
Investments in networks: assumptions I

- $\Gamma$ has an initial round 0 that allows for investments in an information network: actors can establish a network for information exchange between trustors.

- *Without* investments in the network, each trustor $i$ is informed in round $t = 2, 3, ...$ only about the history of her own TGs in all previous rounds $1, ..., t - 1$ with the trustee.

- *With* investments in the network, each trustor $i$ is informed in round $t = 1, 2, ...$ also about the history of the TGs of the other trustor $j$ in all previous rounds $1, ..., t - 1$ and about the history of the TG of the other trustor in round $t$ if $j$ played before $i$ in round $t$.

- *With* and *without* investments in the network, the trustee is always informed about the history of all his TGs with both trustors.

- No noise: information – if available – is correct.
Investments in networks: interpretation

- Note: the network between trustors allows for (1) information exchange and (2) for conditioning own present behavior not only on own previous experience with the trustee but also on previous behavior of the trustee in TGs with other trustors
- Thus: investments in networks increase the trustors’ sanction potential (control opportunities) vis à vis the trustee
- Network as an informal institution, established by the actors themselves, that supports (norms of) cooperation, i.e., supports (norms of) trustful and trustworthy behavior
Investments in networks: example

- Consider the example of a market with two buyers as trustors who purchase goods from a seller as trustee under conditions of asymmetric information.
- The market participants can set up a consumer organization or a buyer association that keeps track of transactions.
- The information network between buyers that provides information about the behavior of the seller is then due to the distribution of information on the behavior of market participants by the organization.
Investments in networks: assumptions II

- Costs of investments in the network: \( \tau > 0 \)
- Two alternative scenarios for sharing the costs of investments:
  - *Trustor scenario*: only the trustors can invest in round 0
  - *Trustee scenario*: only the trustee can invest in round 0
Trustors can invest in networks

Trustor scenario: only the trustors can invest

- In round 0, each trustor decides on whether or not to invest $\tau/2$
  - Trustors decide simultaneously and independently
- If both trustors decide to invest, trustors exchange information in all subsequent rounds
- Otherwise, no information exchange between trustors in subsequent rounds and no trustor loses her investment $\tau/2$

- Note: the assumptions for the trustor scenario emulate the standard assumption on two-sided link formation: both actors have to agree on forming the link and share the costs
Trustee can invest in networks

*Trustee scenario:* only the trustee can invest

- In round 0, the trustee decides on whether or not to invest $\tau$
- If trustee decides to invest, trustors exchange information in all subsequent rounds
- Otherwise, no information exchange between trustors in subsequent rounds

- Note: investment of the trustee as a voluntary “commitment” of the trustee that supports (norms of) cooperation, i.e., supports (norms of) trustful and trustworthy behavior
Further assumptions

● (Expected) payoff for $\Gamma = \text{investments in round 0 + discounted sum of payoffs in rounds 1, 2,...}$

● Structure of the game is common knowledge

● $\Gamma$ is a noncooperative game
Analysis of the model
Approach

- We derive conditions for subgame perfect equilibria (spe) of $\Gamma$ such that trust is placed and honored in all TGs in all rounds ("trust equilibria")
  - Note: actors follow a norm of cooperation in such an equilibrium
- We are specifically interested in spe’s such that trust is placed and honored in all TGs in all rounds if actors have invested in the network in round 0, while trust is never placed (and would be always abused) without investments in round 0
- We first analyze subgame $\Gamma^-$ that is played without investments in round 0 (question: norms of cooperation through repeated interactions?) and subgame $\Gamma^+$ that is played after investments in round 0 (question: norms of cooperation through networks?)
- We then derive conditions such that investments in round 0 are on the spe-path
Subgame perfect equilibrium

- Very roughly: a subgame perfect equilibrium of $\Gamma$ is a combination of strategies such that each actor’s strategy maximizes that actor’s payoff against the other actors’ strategies for each situation that may emerge in $\Gamma$ (including situations off the equilibrium-path)
Trust via conditional strategies I

- One easily verifies that placing trust *unconditionally* is inconsistent with equilibrium behavior (since the trustee would then be better off by abusing trust)

- We derive conditions for an spe such that trust is placed and honored *conditionally* (trust based on “reciprocity”):
  - trustors place trust (and thus reward trustworthy behavior of the trustee) as long as they have no information that trust has ever been abused
  - trustors punish abuse of trust by not placing trust as soon as they have information that trust has been abused
  - trustee honors trust in equilibrium (and abuses trust after deviations from the equilibrium)

- When trustors place trusts conditionally, the trustee must balance the short-term incentive to abuse trust (T – R) and the long-term incentives of not being trusted (R – P)
**Trust via conditional strategies II**

- Most severe punishment for abuse of trust: trust is *never* placed again $\Rightarrow$ trigger strategy
- Conditions for a trust equilibrium with trigger strategies for trustors are necessary and sufficient conditions for spe such that trust is always placed and honored
- Assumption for empirical applications: trust and trustworthiness are more likely when the conditions for a trust equilibrium with trigger strategies are less restrictive
Conditional strategies and norms of cooperation

- Strategies are rules guiding social behavior
- Actors using trigger strategies (or other conditional strategies) thus (1) follow a conjoint social norm of conditional cooperation, (2) sanction others positively by placing and honoring trust who cooperate, (3) sanction others negatively by no longer placing and honoring trust who do not cooperate
- Trust equilibrium $\iff$ following the norm is equilibrium behavior (i.e., the norm is self-enforcing)
- If investments in networks are part of equilibrium behavior, it is in the actors’ self-interest to establish an informal institution that supports the norm of cooperation
Trust through repeated interactions without a network

- **Proposition 1**: $\Gamma^-$ has an spe with trust placed and honored in all TGs in all rounds iff

\[
 w \geq TEMP^- := \frac{T - R}{T - P}
\]

- Note: this is the standard condition for an spe such that trust is placed and honored in an indefinitely repeated TG of a trustee with one trustor. Under this condition, if both trustors play trigger strategies, the trustee’s payoff from abusing trust in both TG’s in round 1 is not larger than his payoff from always honoring trust.
Trust in a network

- **Proposition 2:** $\Gamma^+$ has an SPE with trust placed and honored in all TGs in all rounds iff

$$w \geq TEMP^+ = \frac{T - R}{(T - P) + (R - P)}$$

- **Sketch of proof:** If both trustors play trigger strategies and the trustee has a better strategy than always honoring trust, he also benefits from honoring trust in the first TG in round 1 and abusing trust in the second TG in round 1. Under the condition, the payoff from such a “deviation” is not larger than the payoff from always honoring trust.
Networks and norms of cooperation

- Properties of $TEMP$:
  \[
  0 < TEMP^+ < TEMP^- < 1
  \]

- The threshold $TEMP$ for the continuation probability $w$ such that an SPE exists with trust placed and honored in all TGs in all rounds is smaller with a network (in $\Gamma^+$) than without a network (in $\Gamma^-$).

  ➔ The conditions for a self-enforcing norm of cooperation in the TG are less restrictive with a network than without a network, i.e., networks can support norms of cooperation in social dilemmas.

- This is due to the fact that the information network between trustors increases their sanction potential (control opportunities) vis-à-vis the trustee.
The value of the network

*Proposition 3* – Assume that an equilibrium with trust always placed and honored exists in $\Gamma^+$ but not in $\Gamma^-$, i.e.

$$ \text{TEMP}^+ \leq w < \text{TEMP}^- $$

The upper bound on the value of the network is then

$$ \frac{2(R - P)}{1 - w} $$

for the trustee and

$$ \frac{R_i - P_i}{1 - w} $$

for each trustor $i$
Network formation I

- Proposition 4 – Investments of the trustors: \( \Gamma \) has an SPE in the trustor scenario such that the trustors invest in the network and subsequently trust is always placed and honored if

\[
(1) \quad TEMP^+ \leq w < TEMP^-
\]

and

\[
(2) \text{ for each trustor } i: \quad \frac{\tau}{2} \leq \frac{R_i - P_i}{1 - w}
\]
Network formation II

- Proposition 5 – Investments of the trustee: \( \Gamma \) has a spe in the trustee scenario such that the trustee invests in the network and subsequently trust is always placed and honored if

\[
(1) \quad TEMP^+ \leq w < TEMP^-
\]

and

\[
(2) \quad \tau \leq \frac{2(R - P)}{1 - w}
\]
Testable predictions: network formation

- Propositions 4 and 5 imply that rational actors invest in networks if trust problems are “intermediate”, i.e., neither too small (so that rational trust is possible without networks) nor too large (so that rational trust is impossible even with investments in networks), and if the costs of investments are small enough
  - Investments are more likely if trust problem is “intermediate”
  - Costs of investments have stronger effect on investment behavior if trust problem is “intermediate”
Testable predictions: effects of repeated interactions and networks

- Predictions on effects of repeated interaction and network effects are predictions on (1) behavior of trustors and trustee in the TGs after round 0 and on (2) the effects of behavior in round 0 on behavior in subsequent TGs
  - These predictions basically depend on whether or not $\Gamma^+$ and $\Gamma^-$ have trust equilibria
  - Predictions are similar to those from standard models of indefinitely often repeated games and models of network effects
- Basically: (1) likelihood of following a norm of cooperation (i.e., placing and honoring trust) increases in $w$ and decreases in $\text{TEMP}$; (2) effects of behavior in round 0 and network effects on following a norm of cooperation (i.e., placing and honoring trust) are large for “intermediate” trust problems

Details: Raub, Buskens & Frey 2013
Summary and conclusions
Summary I

- "Demand" for norms of cooperation in social dilemmas: each actor is better off when everybody cooperates than when everyone defects, while there are individual incentives to defect.

- "Realization" of norms: in repeated social dilemmas, following a norm of cooperation can be in the actors’ “enlightened” self-interest in the sense that following the norm is equilibrium behavior (*self-enforcing* norms).

- The norms of cooperation prescribe to place and honor trust conditionally (“reciprocity”).
Summary II

● Networks can support norms of cooperation in social dilemmas and can reinforce the effects of repeated interactions through information exchange, thus increasing actors’ sanction potential

● Hence, actors have incentives to invest in networks → strategic network formation

● New contribution: integrated model of network formation and effects of networks

  ▪ We assume full strategic rationality, also with respect to network formation

● Model yields testable implications on network formation and effects of repeated interactions and networks on norms of cooperation
Generalizations and extension

- Very similar results are obtained for (see Raub, Buskens & Frey 2013)
  - A model for more than 2 trustors
  - A model for more than 1 trustee
  - Other 2-person social dilemma games: Investment Game, Prisoner’s Dilemma, two-actor Public Goods Game, and others

- Extension: modeling network formation and effects of networks in trust problems with incomplete information (Frey, Buskens & Raub 2015)
  - Generalizations of that model for other social dilemmas than the Trust Game are, however, far from trivial
Strengths and weaknesses of the model; next steps

- The model shows that rational actors can, in principle, realize norms of cooperation in social dilemmas.
- But: folk theorem and equilibrium selection problem: (many) other equilibria of repeated games.
- Hence: the model contributes to understanding the maintenance of norms of cooperation (in the sense of showing that actors have no incentives to deviate from the norm, once the norm has been established and everyone expects everybody else to follow the norm) but much less – if anything – to the emergence of such norms (in the sense of contributing to the understanding why a norm of cooperation has emerged rather than some other rule of behavior that is supported by an equilibrium).
- Also: what about norms of cooperation in non-repeated (“one-shot”) social dilemmas? ➔ topic of next part of the workshop.
References

Appendix
Further limitations of the model

- If trustors differ markedly in $R_i - P_i$ (the cost of distrust relative to honored trust), the assumption that they share investment costs equally might not be plausible.
- We neglect scenarios with trust in some but not in all TGs.
- We assume a game with complete information, hence:
  - the network is not used for learning about unobservable characteristics of the trustee.
  - trust depends exclusively on incentives for the trustee.
- We assume away problems (noise or incentive problems) related to the diffusion of information.
- We use assumptions that reduce collective good problems associated with investments in social capital.
A “dilemma” with respect to game theoretically inspired models of network formation and network effects I

- **New feature** and (in a sense) **strength** of the model: integrated model of network formation and network effects, assuming full strategic rationality

- **Weaknesses** of the model:
  - No account for network structure: analysis is restricted to the extreme cases of an empty or complete network
  - No account for network dynamics: the network is formed once and for all at the beginning of the game
  - Hardly any account for actor heterogeneity: allowing for heterogeneity with respect to actors’ payoffs in the underlying social dilemma requires restrictive assumptions on the order of play in rounds $t = 1, 2,...$
A “dilemma” with respect to game theoretically inspired models on network formation and network effects II

● “Dilemma”: accounting for network structure, network dynamics, and actor heterogeneity is typically unfeasible under the assumption of full strategic rationality

● Accounting for such features typically requires “ad hoc” assumptions on regularities of human behavior (e.g., “bounded rationality” assumptions such as the assumption of myopic best-reply behavior in network formation models)
  ▪ These assumptions are typically not (or at least not easily and convincingly) applicable in other contexts

● Again the question arises: what are alternatives to standard rational choice models that are tractable, as broadly applicable, and at least as well corroborated empirically as standard rational choice assumptions?
Summary

- Integrated model of network formation and effects of networks
  - Model with full strategic rationality, also with respect to network formation
- Model yields (testable) implications on network formation and effects of networks
- Network between trustors allows for (1) information exchange and (2) sanctioning opportunities
- “Value of the network” is an implication of the model and is precisely defined
- Note: investment of the trustee as a voluntary “commitment” that supports (norms of) cooperation, i.e. supports (norms of) trustful and trustworthy behavior
Social norms of cooperation through endogenous peer punishment
Social norms in one-shot games

- Public Goods Game, n-person PD
- Less focus on game-theoretic arguments
- More focus on experimental findings
- Norm enforcement through additional moves in the games: sanctioning institutions rather than repeated games and networks
- Again: attention for endogenous formation of institutions
Social norms in one-shot games

- Standard game-theoretic prediction:
  - No contribution in Public Goods Games
  - No cooperation in PD
  - Prediction does not change with introducing costly sanctioning options

- Observed behavior: considerable amount of contributions/cooperation, at least initially

- Sanctions can sustain/promote a norm of cooperation
Example of Fehr and Gächter

(Nature 415, 137-140, 2002)
Game-theoretic explanation

- Actors have different preferences and some have non-standard preferences
- Some actors want to contribute if enough others contribute as well
- Some want to bear costs to sanction others who do not follow a cooperative norm
- Often the subjects who want to contribute if others contribute are the same as those willing to bear costs to sanction non-contributing others
Behavior in the PGG (Fischbacher et al. 2001)

Empirical evidence for heterogeneous preferences
Most people reciprocate expected average contribution of others ('conditional cooperators')
Conditions to maintain a norm of cooperation

- Conditions need to be favorable enough to solve the first as well as the second order free-rider problem
  - Enough contributors
  - Enough people willing to punish free-riders
  - Not too many people punishing cooperative others
Recent work on the design of sanctioning institutions

- Punishments or rewards (Van Lange et al. 2014)
- Size of the stick (Egas and Riedl 2008)
- Individual vs collective sanctions (Van Miltenburg et al. 2014)
  - Avoiding punishment of cooperative others
- What if behavior cannot be observed perfectly (noise)?
- Which institutions do actors choose themselves?
Endogenous institutional choice

• Do actors want peer punishment?
• How large is the stick they choose?
• Does this depend on whether contributions can be observed perfectly or not?

• Question: do actors anticipate opportunities of a peer punishment institution?
• Previous research: most actors prefer peer punishment institution after some experience
Gürerk et al., 2006, Science
An experiment on 6-person PD games with noise and endogenous peer punishment
The interaction situation

- Groups of 6 actors play series of one-shot PDs
- Actors receive an endowment of 20 points
- Only full contribution or no contribution can be chosen
- Multiplier for contributed amount is 2.4
- Two variants:
  - No noise: contribution decisions of other group members are perfectly visible
  - Noise: when group members observe each other’s contribution decision, there is a 20% independent chance that they observe a contribution as a defection and vice versa
The voting stage

- Before each PD, all group members participate in a “voting stage”
- Two simultaneous votes:
  - Add a punishment stage to the PD: yes or no?
  - If a punishment stage is added, small or large stick?
- When they vote, actors are aware of noise with respect to contributions
- If a majority votes for a punishment stage, this stage is added to the PD with the size of the stick that the majority of the group voted for
  - Ties are broken randomly
The interaction situation

● A punishment stage implies that after observing the (noisy) contributions every group member can decide whether or not to punish every other
  ▪ Cost of allocating a punishment: fixed at 2 points
  ▪ Payoff-reduction recipient is endogenous
    ▪ Small stick: 6 points (2:6)
    ▪ Large stick: 12 points (2:12)
● Note: given the behavior of everyone else payoff defect − payoff contribute = 12 points
Theoretical framework

- Without a punishment institution, actors earn most by defecting; actors with “standard” preferences will drive out cooperation.
- Under a punishment institution: actors with “standard” preferences might contribute if expected punishment is large enough, depending on noise and size of the stick.
- Expected overall benefits of the punishment institution depend on the number of actors willing to punish and the number of actors with “standard” preferences willing to contribute under the punishment institutions.
Main arguments

- Under noise, punishment will be more often directed at contributors, and more defectors will go unpunished.
- Moreover, even if one cooperates, one might get punished because of a wrong observation by someone else.
- Thus, punishment institutions are less effective and larger sticks are a risk to oneself as well.
- The following hypotheses follow from these arguments, but also from more formal arguments based on the assumption that some actors in the population have a preference for reciprocation.
Hypotheses

● Subjects vote less often for a punishment institution under noise than without noise
● Subjects vote less often for a large stick under noise than without noise
● Over time, choice of institutions depends on their success:
  ▪ More cooperation under a punishment institution will lead to a recurrent choice of this institution
  ▪ More cooperation with no punishment institution will make choosing a punishment institution less likely
Experiment

- 156 subjects interact in series of one-shot 6-player PDs
- Two between-subject conditions: with and without noise
- Every session starts with 5 periods without voting and punishment stage
- Followed by 40 periods with voting stage, endogenous punishment stage, and endogenous size of stick
- Subjects are informed about (noisy) contributions and own (noisy) earnings
- If a punishment stage is implemented, subjects are informed about all allocated punishments in their group
Results – votes for institution

- No noise

- Noise
Results – votes for institution

- Peer punishment less often voted for under noise
- Also, big sticks less often voted for under noise

- Without noise: peer punishment and especially big sticks gain popularity over time
- With noise: only small sticks are chosen occasionally, big sticks do not gain ground over time
Results – contributions

● No noise

![No noise graph](image)

● Noise

![Noise graph](image)
Results – contributions

● Without noise: contributions are very high with both punishment institutions
  ▪ Subjects quickly learn that these institutions can promote contribution and then they also choose them
  ▪ Executing punishment is hardly necessary
● With noise: contributions are also very high with punishment institutions
  ▪ Subjects have more trouble to learn, because they will always observe some free-riding
  ▪ They become much less likely to support the punishment institutions also because they observe punishment of cooperators
Results – earnings under voting outcomes

- No noise

![Graph showing average earnings over periods with voting rules for no pun, low effort, and high effort cases with and without noise.]

- Noise

![Graph showing average earnings over periods with voting rules for no pun, low effort, and high effort cases with noise.]
Results – actual and observed cooperation

![Graph showing predicted probability to contribute and observe a cooperator with and without noise.](image-url)
Results – actual and observed earnings

![Graph showing predicted and observed period earnings with and without noise.]

The graph illustrates the predicted and observed period earnings with and without noise. The earnings are categorized into NP, LP, and HP levels. The bars on the left represent the earnings without noise, while the bars on the right show the earnings with noise. The graph indicates a significant difference in earnings with the presence of noise compared to the no noise condition.
Conclusions

- Confirmation of previous findings for games without noise:
  - Most groups implement a peer punishment institution after some experience
  - Most groups allow severe punishments
- However, completely different results with noise:
  - Most groups interact without a punishment institution
  - Very reluctant to allow severe punishments
  - Observed earnings are lower with than without a punishment institution
  - Observing that cooperators are punished discourages actors from voting in favor of punishment institutions
Conclusions

● In one-shot interaction without cooperative equilibria under standard game-theoretic assumptions, actors are able to self-organize (with endogenous punishment institutions) in cooperative groups with a punishment institution in place.

● However, when cooperative behavior is not easily observable, a punishment institution can still support a norm for cooperative behavior, but it is much less likely that actors do self-organize into an institution with peer punishment.
This study and related studies, see PhD thesis Nynke van Miltenburg

- Collective vs individual peer sanctions
- Fairness of peer sanctions
- Noisy contribution decisions
- Endogenous institutions
- Study presented co-authored by N. van Miltenburg, V. Buskens & W. Raub
Overall conclusions
Game theory and norms I

- Game theory provides theoretical tools to help us understand:
  - The emergence of norms
  - The maintenance of norms
  - The emergence and maintenance of institutions to sanction behavior deviant from a norm
- But: game-theoretic equilibria are neither sufficient nor necessary for explaining the emergence and maintenance of norms
Game theory and norms II

- Multiplicity of equilibria such as in Volunteers Dilemmas and repeated Trust Games leaves us with questions on which norm emerges under which conditions.
- For repeated social dilemmas, necessary and sufficient conditions for norms of cooperation can be derived and can be used for predictions under which conditions norms of cooperation will be maintained.
- For one-shot games without cooperative equilibria under standard conditions, sufficient conditions in terms of the existence of people with non-standard preferences can be derived that allow for predictions of conditions favorable for the emergence and maintenance of norms of cooperation.
Game theory and norms III

- Experimental research can be used to:
  - Test the likelihood of emergence of cooperative norms under different conditions / institutions
  - Investigate the conditions under which institutions to promote cooperative norms emerge themselves and provide more evidence on the conditions under which it is more or less likely that groups self-organize and are able to maintain cooperative behavior
References