Sanction recognition: A simulation model of extended normative reasoning

Martin Neumann
Ulf Lotzmann
University of Koblenz
Overview

1. Introduction / Context
2. Norms in the crime scene
3. Theoretical background
4. Empirical background
5. Conceptual model: *Problem*
6. Intra agent processes &
7. (preliminary) Simulation results: *Solution*
Context

• Context: Part of EU FP 7 Project GLODERS
  – Simulation of Extortion Racket Systems
  – Investigation of two kind of norms:
    • Investigating norms regulating relations *between* civil society and criminal organizations (fight against extortion racket systems)
    • Investigating *intra-organizational norms* in criminal organizations (violence between criminals)
  – Objective of this talk: *intra-organizational norms* in the crime field
  → Case study: discovery of norm enforcement as theoretical problem
Norms in the crime scene

• Characteristics of Criminal organizations:
  – Emerge (often) from the scratch for making use of criminal opportunities
  – Need to deal with situations as they arrive
    • Characteristic: Fluidity of relations
  – Norms not codified: emerge during interactions
    • Characteristic: Ambiguity of norms
  – No recourse to court as third party for conflict resolution
    • Typical norm (covertness): don’t talk too much
      → Therefore information unreliable
Theoretical background

• Norm theories: *role theory*
  – Internalization: no theory of norm enforcement

• Norm theories: *social identity*
  – Role model small groups. Norm enforcement by learning. Problem: large anonymous societies

• Norm theories: *rational choice*
  – Theory of norm enforcement: sanctions
  – Sanctions: basic principle
Theoretical background

• Rational choice: Theory of individual decision making
  • Norms: behavior regularity in populations
  • Mechanism of norm enforcement: sanctioning norm deviation
    – Game theory
      • Extending Rational choice of action by investigating strategic interaction
      • Cooperation norms
    – Evolutionary Game theory (e.g. Axelrod, 1986)
      • Extending game theory of dynamic perspective
      • Three possible actions: cooperate, defect, sanction
        – Norm = cooperate
        – Aggression = sanction (loss of numerical values)
Theoretical background

– Fundamental insights
  • Evolution of solidarity (De Landa): spreading of strategies in populations (e.g. Axelrod, Nowak)
  • Norms: might push actors from mutual defection (e.g. Ulmann Margalith)
  • Meta-norms (Axelrod)
  • Homo socialis: reproductive success of ‘friendly’ agent by increased cognition (Helbing)

→ Basic theoretical term: sanctions
→ Sanction recognition not discussed
Empirical background

• Empirical case study: analysis of police interrogations

• Empirical case demonstrates:
  – Sanction recognition (empirically) not trivial
  – Error prone

• Due to specific situations of crime scene:
  – Ambiguity
  – Fluidity
Empirical background

• Case: Gang of drug dealers, established ca. 1990

• Collapse ca. 2002 – 2005
  – Series of murders
  – Run on the bank: extortion among criminals
  → Cascading effect of spread of mistrust
    • Mistrust could not be encapsulated
    • Reason: errors in reasoning about aggression
Empirical background

• “emergent” macro-effect (Collapse):
  – Rule of terror in which old friends are killing each other (in vivo-code)
  – Corrupt chaos (in vivo-code)

• Responsibility for participants not visible
  – No individual ruler (Nero burning Rome), but ruled by invisible hand
Conceptual model

• 5 phases can be identified in the dynamics of the collapse
  – Ordinary business: money laundering
  – Crystallising kernel of mistrust
  – Conflict escalation
  – ‘Corrupt chaos’
  – Run on the bank
  • The ‘corrupt chaos’ and the ‘run on the bank’ happened in parallel
Conceptual model

• Crystallising kernel of mistrust
  – Becoming disreputable
    • External event such as someone becoming too greedy
  – Aggression:
    • An attack to the life of M. (survived)
  – Interpreting aggression:
    • M. told the newspapers ‘about my [V01] role in the network’ because he thought that I wanted to kill him to get the money.
Conceptual model

- Reasoning about aggression:
  - Attack against life:
    - Sanction (Death penalty: eventually accused for stealing drugs)
    - Self interest (get his money)
  → Interpretation
Conceptual model

• Reasoning about aggression:
  – Interpretation M:
    • Self interest (V01 want to get his money)
    • Factually wrong that V01 mandated the assassination

Feedback loop
Conceptual model

- Reasoning about aggression:
  - reaction:
    - Counter-aggression (betrayal of role in network)
    - Triggers reasoning about aggression

Feedback loop
Conceptual Model

– Theoretical conclusion:
  • Norm enforcement (sanction) not delegated to specialised institution

– In behavioural terms: Sanction = aggression

– (at least) two possible motivations for aggression
  • Self interest: aggression = norm violation
  • Sanctioning: aggression = Norm enforcement

→ Sanction recognition:
  → Interpretation: necessity of reasoning on aggression
  → Cannot be treated as basic (unexplained) term
Intra-agent processes

• Core: normative process
  – *Norms (very broad):*
    1. No aggression between fellows
       – Violent behaviour
       – Embezzle money
       – making group public
    2. Sanction observed norm deviation
       – Violent behaviour
  → Potential conflict between 1. and 2.
Intra-agent processes

- Overview of model
  - Reaction &
  - Interpretation of aggression
Intra-agent processes

• Characteristics of the agents (aggressor) involved in interpretation of aggression:
  – *Reputation*: Static, objective
  – *Image*: dynamic, subjective
    • Image change due to:
      – Deviant behaviour (decrease)
      – Acceptance of punishment (Obedience): image of victim of aggression (deviant agent) increases
      – Perception of aggression as norm enforcement: image of aggressor increases
      – Perception of aggression as norm violation: image of aggressor decreases
Intra-agent processes

1. Decision about objective characteristic of aggressor: reputable?
2. Decision about aggression: norm violation?
3. Decision about subjective characteristic of aggressor: image?
4. Decision about reaction: obedience?

---

Agent X: Interpretation of aggression

- **Interpret as possible sanction**
  - C1: attacker reputation (very) high
  - C4: no norm violation found
  - C3: possible norm violation with temporal relation to reproach found
  - Obey
- **Interpret as possible rebuke**
  - C2: attacker reputation modest or (very) low
  - C5: attacker image (very) high and mild or modest aggression
  - C6: attacker image (very) low or modest or strong aggression
- **Interpret as unjust aggression**
  - C7: strong or modest violence or strong thread
  - Switch to emotional frame

---

Increase image of aggressor
Decrease image of aggressor
Simulation experiments
Simulation experiments

• Initial loss of trust against one member
• Observed by reputable criminal
• Decision for punishment: ‘outburst of rage’
Simulation experiments

• Aggressor is reputable
• Possibly justified aggression
• Victim of aggression cannot find norm violation
Simulation experiments

• However, obedience due to high image of aggressor
(preliminary) Simulation results

<table>
<thead>
<tr>
<th>Tick</th>
<th>Criminal-6</th>
<th>Reputable Criminal-0</th>
<th>Reputable Criminal-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>modest</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>modest</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>3</td>
<td>modest</td>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>4</td>
<td>modest</td>
<td>high</td>
<td>veryhigh (sanction)</td>
</tr>
<tr>
<td>5</td>
<td>modest</td>
<td>high</td>
<td>veryhigh</td>
</tr>
<tr>
<td>6</td>
<td>modest</td>
<td>high</td>
<td>veryhigh</td>
</tr>
<tr>
<td>7</td>
<td>modest</td>
<td>high</td>
<td>veryhigh</td>
</tr>
<tr>
<td>8</td>
<td>modest</td>
<td>high</td>
<td>veryhigh</td>
</tr>
</tbody>
</table>

Excerpt of image of Criminal 1

- **Development of Normative Authority**
  - Temporary
  - Subjective

Need not be stable: *fluidity*

Need not converge between agents: *Ambiguity*
Thank you for your attention