Input (rules specifications) for PROTON simulations and Wizard
Deliverable 4.1
Work Package 4 (Tasks T 4.1, T 4.2, T 4.3)

March 2019

Lead beneficiary: HUJI
Other beneficiaries UCSC, Fraunhofer, ITTI, CNR
Technical References

**Project Acronym** | PROTON
---|---
**Project Title** | Modelling the PRocesses leading to Organised crime and TerrOrist Networks
**Project Coordinator** | Ernesto Savona  
Università Cattolica del Sacro Cuore  
ernesto.savona@unicatt.it
**transcrime**  
Joint Research Centre on Transnational Crime
**Project Duration** | October 2016 – September 2019 (36 months)

| Deliverable No. | D4.1
| Dissemination level | PU
| Work Package | WP 4 – Input Selection & Integration with Experiments
| Task | T 4.1, T 4.2, T4.3
| Lead beneficiary | 2- HUJI
| Contributing beneficiary(ies) | 1 (UCSC), 2(HUJI), 3 (Franhofer), 5 (ITTI), 6 (CNR)
| Due date of deliverable | 31 March 2019
| Actual submission date | 28 March 2019

1 PU = Public  
PP = Restricted to other programme participants (including the Commission Services)  
RE = Restricted to a group specified by the consortium (including the Commission Services)  
CO = Confidential, only for members of the consortium (including the Commission Services)
<table>
<thead>
<tr>
<th>V</th>
<th>Date</th>
<th>Beneficiary</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.03.2019</td>
<td>HUJI</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>28.03.2019</td>
<td>UCSC</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10.05.2019</td>
<td>HUJI</td>
<td></td>
</tr>
</tbody>
</table>
Outline of this deliverable

This deliverable presents the findings and reports for all task within WP4. WP4 includes three tasks, each reported under a separate section.

- Section 1 includes the report on the internal workshop conducted in task T4.1.
- Section 2 presents the report on the operationalisation of factors into input for ABM simulations resulting from the activities conducted under task T4.2. Due to the decision of developing separate simulations for organised crime and terrorist networks, Section 2 is divided into two parts:
  - Report on the operationalisation of factors into input for ABM simulations on organised crime
  - Report on the operationalisation of factors into inputs for ABM simulations on radicalization
- Section 3 comprises the reports for selected experiments conducted in task T4.3. Accordingly, it is divided into two parts:
  - An Experiment on the Individual and Strategic Determinants of Criminal Collaboration
  - Examining the Interactive Effects of Personalization Algorithms (the Filter Bubble) on Network Structure (the Echo Chamber) and the Impact on Radical Beliefs
# Table of content

## OUTLINE OF THIS DELIVERABLE

### SECTION 1. REPORT ON THE INTERNAL WORKSHOP CONDUCTED IN TASK T4.1.

1. **Introduction** 8
2. **Minutes of the Internal Meeting aimed to discuss the results from WP1-3 (M17)** 8
3. **Agenda** 8
4. **Minutes, 19 February 2018**
   4.1. List of participants 9
   4.2. Welcome 9
   4.3. First session: 10:00 – 12:45: Discussion of the organised crime recruitment model (Part 1) 10
   4.4. Second session: 14:00 – 15:00: Discussion of the organised crime recruitment model 13
   4.5. Third session: 15:00 – 16:00: Discussion of the organised crime laboratory experiment 14
   4.6. Fourth session: 16:00 – 17:00: Discussion of the on-line experiment on recruitment to terrorist networks 15
5. **Minutes, 20 February 2018**
   5.1. List of participants 17
   5.2. Fifth session: 10:00 – 12:15: Discussion of the organised crime recruitment model 17
   5.3. Sixth session: 13:30 – 17:00: Follow up on the terrorism recruitment model and subsequent steps 19

### SECTION 2. REPORT ON THE OPERATIONALISATION OF FACTORS INTO INPUT FOR ABM SIMULATIONS RESULTING FROM THE ACTIVITIES CONDUCTED UNDER TASK T4.2

**REPORT ON THE OPERATIONALISATION OF FACTORS INTO INPUT FOR ABM SIMULATIONS ON ORGANISED CRIME** 23

1. **Introduction** 23
2. **From the factors for recruitment and criminalization into OC to input for ABM** 23
   2.1. Results of the systematic review 23
   2.2. Theoretical Framework 25
3. **General structure of the OC ABM** 26
   3.1. A Multiplex network ABM 27
   3.2. Recruitment into organized crime 29
4. **Data** 33
5. **A draft ODD+D for the model of recruitment into Organised Crime** 35
6. **Policy Scenarios** 51

---

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 699824.
6.1. PRIMARY AND SECONDARY SOCIALISATION SCENARIO 51
6.2. LAW ENFORCEMENT SCENARIO 53
7. REFERENCES 54

REPORT ON THE OPERATIONALISATION OF FACTORS INTO INPUTS FOR ABM SIMULATIONS ON RADICALIZATION 57

1. INTRODUCTION TERRORIST NETWORK 57
2. RESULTS OF THE SYSTEMATIC REVIEW 58
3. THEORETICAL FRAMEWORK 58
4. GENERAL STRUCTURE OF THE ABM 61
  4.1. COMMUNITIES 63
  4.2. SETTING UP LOCATIONS 64
  4.3. SETTING UP CITIZENS AND THEIR ATTRIBUTES 64
  4.4. ACTIVITIES 66
5. INPUTS 68
6. DATA 70
  6.1. SETTING: NEUKÖLLN 71
  6.2. DATA 72
7. POLICY SCENARIOS 73
  7.1. EMPLOYMENT 73
  7.2. COMMUNITY CENTER WORKERS 74
  7.3. COMMUNITY POLICING 75
8. REFERENCES 76

SECTION 3 REPORTS ON EXPERIMENTS (TASK T4.3) 80

AN EXPERIMENT ON THE INDIVIDUAL AND STRATEGIC DETERMINANTS OF CRIMINAL COLLABORATION 81

1. SUMMARY 81
2. INTRODUCTION 81
3. STRUCTURE AND RESEARCH QUESTIONS 83
4. EXPERIMENTAL PROCEDURE 84
  4.1. START OF THE EXPERIMENT 84
  4.2. PHASE 1: INDIVIDUAL DIE-ROLL TASK 85
  4.3. PHASE 2: SENDER-RECEIVER TASK 87
  4.4. PHASE 3: COLLABORATIVE DIE-ROLLING 92
  4.5. QUESTIONNAIRE 94
  4.6. TREATMENTS AND SESSIONS 94
5. RESULTS 94
  5.1. OPTING INTO COLLABORATION 96
  5.2. REPORTED ROLLS WHEN COLLABORATING 100
  5.3. BEHAVIOUR OF SECOND MOVERS 104
6. CONCLUSIONS 109
7. REFERENCES 111
EXAMINING THE INTERACTIVE EFFECTS OF PERSONALIZATION ALGORITHMS (THE FILTER BUBBLE) ON NETWORK STRUCTURE (THE ECHO CHAMBER) AND THE IMPACT ON RADICAL BELIEFS 112

1. INTRODUCTION 112
2. THEORETICAL FRAMEWORK 114
   2.1. RADICALIZATION AND ONLINE RADICALIZATION 114
   2.2. THE ECHO CHAMBER: THE IMPORTANCE OF NETWORK LEVEL CHARACTERISTICS 116
   2.3. THE FILTER BUBBLE 117
3. THE CURRENT STUDY 119
   3.1. TREATMENT AND SETTING 120
4. SAMPLE 121
5. ANALYTIC STRATEGY 122
6. VARIABLES AND HYPOTHESES 123
   6.1. DEPENDENT VARIABLE 123
   6.2. INDEPENDENT VARIABLES 124
7. RESULTS 126
8. DISCUSSION 127
9. RELEVANCE TO AGENT BASED MODEL 131
10. CONCLUSIONS 132
11. REFERENCES 133
Section 1. Report on the internal workshop conducted in task T4.1.

Authors: CNR, LUISS

1. Introduction

This document includes a report and a home video recorded during the internal workshop (https://www.projectproton.eu/filming-birth-idea/) - organized by ISTC-CNR - held at LUISS University (Rome) on the 19th and 20th of February 2018. The workshop was aimed to discuss the results of WP1-3 and the organisation of work leading to the PROTON-S simulator and PROTON Wizard.

2. Minutes of the Internal Meeting aimed to discuss the results from WP1-3 (M17)

February 19th and 20th 2018

The meeting—organized by ISTC-CNR—was held at LUISS University Campus, viale Romania, 32, Rome, on the 19th and 20th of February 2018. The meeting was aimed to discuss the results of WP1-3 and the organisation of work leading to the PROTON-S simulator and PROTON Wizard.

3. Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>Agenda Item</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monday, February 19th</strong></td>
<td></td>
</tr>
<tr>
<td>10:00 - 12:45</td>
<td>Discussion of the organised crime recruitment model (part 1)</td>
</tr>
<tr>
<td>14:00 - 15:00</td>
<td>Discussion of the organised crime recruitment model (part 2)</td>
</tr>
<tr>
<td>15:00 - 16:00</td>
<td>Discussion of the laboratory experiments on organised crime</td>
</tr>
<tr>
<td>16:00 - 17:00</td>
<td>Discussion of the on-line experiments on terrorism</td>
</tr>
<tr>
<td><strong>Tuesday, February 20th</strong></td>
<td></td>
</tr>
<tr>
<td>10:00 - 12:15</td>
<td>Discussion of the organised crime recruitment model (part 3)</td>
</tr>
<tr>
<td>13:30 - 17:00</td>
<td>Follow up on the terrorism recruitment model and subsequent steps</td>
</tr>
</tbody>
</table>
4. Minutes, 19 February 2018

4.1. LIST OF PARTICIPANTS

Giulia Missikoff Andrighetto ISTC-CNR
Mario Paolucci ISTC-CNR
Aron Szekely ISTC-CNR
Nicolas Payette ISTC-CNR
Paola Trussardi ISTC-CNR
Giulia Bonelli ISTC-CNR
Vito Trianni ISTC-CNR
Ernesto Ugo Savona UCSC
Francesco Calderoni UCSC
Gian Maria Campedelli UCSC
David L. Weisburd HUJI
Michael Wolfowicz HUJI
Badi Hasisi HUJI
Alese Wooditch Temple University
Daniela Di Cagno LUISS
Werner Güth LUISS
Andrej Angelovski LUISS
Francesca Marazzi LUISS
Andrea Mario Lavezzi UNIPA

4.2. WELCOME

Dr. Giulia Andrighetto (ISTC-CNR)
Session aimed to discuss the main structure, features and requirements of the organised crime recruitment model.

**Prof. Francesco Calderoni (UCSC-Transcrime) - Presentation of the ODD+D document for the Organised Crime recruitment model**

Before the meeting, UCSC-Transcrime had prepared a document for the organised crime recruitment model, following the Overview, Design Concepts and Details (ODD+D) protocol. This document had been discussed during two virtual meetings with ISTC and UNIPA (2018-01-21; 2018-02-05) and in this session there was a detailed discussion of the document, led by Prof. Calderoni.

Prof. Calderoni gives a general presentation of the proposed model of recruitment to organised crime. The model is to be grounded in differential association theory and makes extensive use of the concept of multiplex networks. Multiplex networks can represent multiple types of links. Four types of links will be represented in the model: *family* links, *friendship* links, *professional* links, and *criminal* links. It is believed that relative position of agents in these networks, in addition to individual characteristics, play an important role in criminal involvement.

Furthermore, as explained by Prof. Ernesto Savona (UCSC-Transcrime), these types of links correspond to different levels of socialisation: *primary socialisation* (family links), *secondary socialisation* (friendship links and professional links) and *criminal socialisation* (criminal links). Accordingly, the agent-based model will offer opportunities for testing policies acting on each of these levels. Examples of policy that have been proposed include: the removal of children from Mafia family (primary socialisation), education-based measures like opening schools or economic measures like reducing income inequality or unemployment (secondary socialisation), and different law enforcement strategies for network disruption (criminal socialisation).
The partners agreed that 1) the model of recruitment to organised crime would be a multiplex network-oriented Agent Based Model and 2) policy interventions to be tested with the Agent Based Model would target primary, secondary and criminal socialisation.

Following this general presentation the scale of the model in terms of population was discussed. Ideally, the model would be able to represent a mid-size European city (for example, the city of Palermo in Italy), but computational constraints may end up limiting the scale. Those constraints will not be fully known until the design of the model is more advanced. Modelling a subset of a city (for example, central Palermo) would also be an option. In any case, sensitivity analysis can be used to assess the impact of the model’s population size on the results.

The temporal scale of the model is discussed next. Prof. David Weisburd (HUJI) questions whether multiple months per tick\(^1\) is an appropriate scale. However, Prof. Savona and Prof. Calderoni point out that some of the policy interventions that could be tested through the model (see below) require time scales that span over years or even generations. Given this requirement, using a very small time in the model would not be computationally feasible. The precise scale to use is left undefined.

Prof. Weisburd remarks that one of the issues with most criminological evaluations of public policies is that they are short-term. Taking advantage of the possibility of assessing long-term policy consequences through agent-based models might be an important outcome of the project.

Prof. Badi Hasisi (HUJI) raises the question of the interaction between the criminal network represented in the model and outside actors and criminals, like family members or associate gangs from

---

\(^1\)A “tick” is a single step of an agent-based simulation, which can represent a given length of real-world time.
different geographic areas. Representing such external ties could allow the evaluation of policies such as border regulations. The idea is left as a potential future extension of the model.

Prof. Mario Lavezzi (UNIPA) raises the question of the role of prisons in the model. The discussion is extended to encompass the representation of institutions in general (schools, for example). Prof. Calderoni suggests that no explicit representation of these institutions as agents is needed and that their effect can be modelled solely through the ties in the network. Which prison or school or other institution an agent is at could be modelled as part of the agent’s internal state.

The role of prisons in the model is discussed further and there is consensus about the fact that imprisonment has a significant impact on the social networks of criminals. The details of how the mechanics of the judicial system should be implemented in the model are left open.

**Decision**

The partners agreed that no explicit geographical representation of institutions like prisons and schools is needed in the model.

One of the proposed policy interventions to be assessed using the agent-based model is the removal of children from mafia families. Prof. Hasisi and Prof. Daniela Di Cagno (LUISS) raise some potential issues regarding the ethics of such an intervention. Prof. Savona points out that at least one judge in Italy has ordered the removal of children from families with strong mafia associations, so the issue is not entirely theoretical and the associated ethical implications make the assessment of potential effects using Agent Based Modelling even more relevant.

Some ideas are proposed regarding which effects of the removal of children from mafia families policy could be tested: 1) the effect of removal children at different ages, 2) the effect of removal compared with that of juvenile detention facilities, 3) the effect that removing one child from a mafia family could have on the siblings of that child, etc. Which of these (or other) effects to test with the model is left as an open question.

**Decision**

The partners agreed that testing the effect of the removal of children from mafia families is a relevant policy question to be assessed with the agent based model.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 699824.
Discussion about the policy of removing of children from mafia families leads to discussion about the mechanisms underlying the effects of the policy. It is believed that network structure, and especially family networks, play a role in the transmission of values. Some values increase the risk of joining organised crime and other values play a protective role against it. This is consistent with the differential association theory literature on which the model is based and it should be possible to draw on this literature to validate the model.

Dr Giulia Andrighetto (ISTC-CNR) points out that modelling the role of values or norms might require the specification of a value/norm-transmission algorithm for the model. ISTC-CNR has expertise in implementing that kind of mechanism, but it is not yet decided if it will be needed for the model. That question will need to be resolved when more progress is made on the model’s design.

Some general discussion about the level of detail that should be included in the model follows. Dr Alese Wooditch (Temple University) argues that only elements that are essential for answering a particular research question should be included. Dr Mario Paolucci (ISTC-CNR), without explicitly endorsing it, mentions another school of thought saying that as many elements as possible should be included at first and then selectively taken out once it is known whether or not they make a difference in the model’s results. Mr. Michael Wolfowicz (HUJI) maintains that whether or not we have empirical data to calibrate and validate the agents’ attributes and their effect should be a most important consideration. Specific attributes proposed for the agents in the model are discussed, but it is too early in the modelling process to make any final decision about particular individual attributes. There is some consensus that the model should start simple and made increasingly more complex if needed.

4.4. SECOND SESSION: 14:00 – 15:00:
DISCUSSION OF THE ORGANISED CRIME RECRUITMENT MODEL

Session aimed to continue the discussion started in the morning about the organised crime recruitment model.
The need to model all four network layers (family, friends, work, and criminals) is reiterated by Prof. Calderoni. Dr Andrighetto reminds the group that each network layer will require formal rules about the way these layers are created and the way they evolve. Some of the questions that will need to be answered for each of the four layers are: How are links created? How are links removed? How are the link weights updated? Dr Andrighetto also points out that there could be interactions between the different layers: interventions on the family network, for example, could also affect the professional network. These interactions will need to be formally specified.

Discussion shifts to possible policy interventions to be explored using the model. One possible area of intervention (in line with the idea of testing interventions impacting secondary socialisation) is to assess measures related to schools and children. Prof. Lavezzi gives a real-world example of organizing football games with children in the city of Palermo, thus getting them off the streets and having them socialize under different conditions. Other ideas include: measures to prevent school dropouts, opening schools during summer, increasing resources for after school activities, programs to retain teachers in difficult schools, etc. Another area of intervention discussed concerns economic measures meant to address income inequality and unemployment. Possible interventions include - for example - government incentives for businesses to hire more people and youth training programs.

While these ideas are deemed by the group to be interesting, no formal commitment has been made to implement any particular one of them yet.

4.5. THIRD SESSION: 15:00 – 16:00: DISCUSSION OF THE ORGANISED CRIME LABORATORY EXPERIMENT

Session aimed to discuss laboratory experiments on recruitment to organised crime.

Prof. Daniela di Cagno (LUISS) - Presentation of the laboratory experiment on recruitment to organised crime

The laboratory experiment on organised crime proposed by LUISS aims to test the relationship between social ties and unethical behaviour. The results of this experiment will fill some of the knowledge-gaps that have not been covered during WP1-3. Additionally, the experiment will com-
plement the organised crime agent-based model because, like the agent-based model, it focuses on the role of social networks.

The experiment will consider dishonesty as the unethical behaviour. This can be measured using well-developed behavioural tools from experimental economics. Specifically, LUISS propose using the ‘die-rolling’ paradigm developed by Fischbacher and Föllmi-Heusi (2013). In this approach, subjects privately roll a die and are asked to report what they rolled; the more they report the more they earn. By comparing the known probabilities with the reported numbers, the experimenters can derive the level of dishonesty of subject groups. LUISS proposes to use a variant of this task in which there are social consequences to dishonesty such that dishonesty affects the earnings of others in a group.

LUISS suggests to test how a series of treatments change the level of dishonesty. They will test the effect of both network structure—whether subjects are linked in a ‘star’ or a ‘circle’—and how the removal of subjects affects dishonesty in other group members and potentially subsequent network formation. The removal of dishonest individuals models in a simple way the effect of imprisonment.

The experiment will be conducted at the Centro Sperimentale Roma Est (CESARE) at LUISS and subjects will be recruited using the system ORSEE and rely on the subject pool of LUISS (primarily a student sample). This provides a base sample upon which to test the effect of networks on this harmful behaviour.

Decision

The partners agreed that Prof. Di Cagno will prepare a document describing the experiments proposed for studying recruitment to organised crime and share it with the other partners, as requested by Prof. Savona.


FOURTH SESSION: 16:00 – 17:00: DISCUSSION OF THE ON-LINE EXPERIMENT ON RECRUITMENT TO TERRORIST NETWORKS

Session aimed to discuss the laboratory experiment on recruitment to terrorist networks.

Mr. Michael Wolfowicz (HUJI) - Presentation of the on-line experiment on recruitment to terrorist networks

The aim of this experiment is to understand the effect of social media on radical beliefs. Building on the research concerning ‘echo chambers’—social networks in which people with similar opinions associate together while they disassociate from those with different opinions—in social media usage, this

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 699824.
experiment will specifically test the role of automatic algorithms.

Its focus will be on the social network Twitter. When users sign-up to join this social network, they are given the opportunity to enable or disable personalisation algorithms. These algorithms change what users are shown based on their behaviour thus personalising the information displayed to them on the network and simultaneously promoting echo chambers.

The experimental treatment will use a between-subjects design and manipulate whether people enable or disable these algorithms during their sign-up procedure. It will then measure, among other factors, the number of radical individuals that experimental subjects are linked with on Twitter after a substantial amount of time has passed (e.g. six months).

Israeli and Palestinian subjects will be recruited who are currently not signed-up to Twitter. This allows HUJI to assist subjects in their sign-up and ask them to enable or disable the personalisation algorithms.

**Decision**

The partners agreed that HUJI proceeds with the experiment.
5. Minutes, 20 February 2018

5.1. List of Participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giulia Missikoff Andrighetto</td>
<td>ISTC-CNR</td>
</tr>
<tr>
<td>Mario Paolucci</td>
<td>ISTC-CNR</td>
</tr>
<tr>
<td>Aron Szekely</td>
<td>ISTC-CNR</td>
</tr>
<tr>
<td>Nicolas Payette</td>
<td>ISTC-CNR</td>
</tr>
<tr>
<td>Paola Trussardi</td>
<td>ISTC-CNR</td>
</tr>
<tr>
<td>Giulia Bonelli</td>
<td>ISTC-CNR</td>
</tr>
<tr>
<td>Francesco Calderoni</td>
<td>UCSC</td>
</tr>
<tr>
<td>Gian Maria Campedelli</td>
<td>UCSC</td>
</tr>
<tr>
<td>David L. Weisburd</td>
<td>HUJI</td>
</tr>
<tr>
<td>Michael Wolfowicz</td>
<td>HUJI</td>
</tr>
<tr>
<td>Badi Hasisi</td>
<td>HUJI</td>
</tr>
<tr>
<td>Alese Wooditch</td>
<td>Temple University</td>
</tr>
<tr>
<td>Daniela Di Cagno</td>
<td>LUISS</td>
</tr>
<tr>
<td>Werner Güth</td>
<td>LUISS</td>
</tr>
<tr>
<td>Andrej Angelovski</td>
<td>LUISS</td>
</tr>
<tr>
<td>Francesca Marazzi</td>
<td>LUISS</td>
</tr>
<tr>
<td>Andrea Mario Lavezzi</td>
<td>UNIPA</td>
</tr>
</tbody>
</table>

5.2. Fifth Session: 10:00 – 12:15: Discussion of the Organised Crime Recruitment Model

Session aimed to continue the discussion started in the previous day about the organised crime recruitment model.

Prof. Francesco Calderoni (UCSC-Transcrime) - Presentation of the ODD+D document, continued

Prof. Calderoni leads by reiterating the need to model the four different types of links involved in the multiplex network (family, friends, colleagues, and criminals). The first three types of links can be informed by research on social networks (e.g. small world networks, homophily, etc.). A more difficult challenge will be to obtain sufficient information for modelling co-offending, criminal networks. Dr
Nicolas Payette (ISTC–CNR) argues that co-offending networks should emerge from the model and that we can then use the information we find on co-offending network to validate the model.

Discussion shifts to the operational definition of “joining an organised crime network” that should be used in the model. Prof. Calderoni claims that legal definitions of organised crime vary from one country to another, but often revolve around the idea of multiple people who commit criminal acts together in a stable manner repeatedly with the same people. Prof. Calderoni and Dr Payette argue that a definition along those lines is sufficient to operationalise the notion of “joining” in the model, by making it a function of the co-offending network. The number of people that must associate together in such a way to constitute an organised criminal group can be left as a parameter of the model, thus allowing the model to represent different legal definitions.

**Decision**

The partners agreed that in the model the notion of agents having joined an organised crime network will be based on their co-offending network.

Prof. Werner Güth (LUISS) suggests that it might be important to take into account the economic impact of criminal activity on agents that are not themselves criminals. It is generally agreed that this could be an important factor, but in the interest of starting with a simple model, it is left as a potential future addition.

Dr Payette raises the question of the model’s level of geographical realism. The model, as described in the ODD+D document, is mainly network-based but does include four distinct geographical regions meant to represent areas of a city having different socio-economic characteristics. Prof. Calderoni explains that another important function of the geographical regions is to introduce some clustering in the networks. Dr Payette is worried about the computational complexity introduced by geography and suggest that socio-economic differences and network clustering could be achieved with other methods.

**Decision**

The partners agreed that the first prototype implementation of the model will not include geography. Geography could be added in a subsequent version if deemed necessary to achieve socio-economic and network clustering.

**Decision**

The partners agreed that UCSC will prepare an updated version of the ODD+D document, including inputs and comments discussed during the meeting. Furthermore, ISTC–CNR will start working on a prototype implementation. Once these two tasks have progressed enough, UCSC and ISTC–CNR will meet to work on the model together.
5.3. **SIXTH SESSION: 13:30 – 17:00: FOLLOW UP ON THE TERRORISM RECRUITMENT MODEL AND SUBSEQUENT STEPS**

Session aimed to continue the progress made in prior meetings on the terrorist recruitment model.

**Mr. Michael Wolfowicz (HUJI) - Presentation of the terrorism recruitment model**

Mr. Michael Wolfowicz reports on the progress made by HUJI since the production by HUJI of the ODD+D document describing the requirements for the model of recruitment to terrorism and the reception of the “Preliminary Terrorism Recruitment Model Description” (document circulated by ISTC-CNR on 2017/09/29).

A summary of the model and its the main parameters are first presented. The *scale of the model* is the first item discussed. HUJI’s original ODD+D document suggested a model of twenty communities containing 40,000 residents in total, but Prof. Weisburd now suggests that four communities, with 20,000 residents in total, might be sufficient. Dr Payette states that the reduction in computational cost resulting from modelling a smaller population is highly desirable. He also points out that the chosen number of communities should be a square number in order to be represented on a square grid (which is the case with four communities).

**Decision**

The partners agreed that the model will aim to include four communities with a total number of twenty thousand residents.

The question of which *city or neighbourhood* to use for calibrating the model is addressed next. Previous discussions between the partners had identified Dutch cities, like The Hague, as potential targets for modelling because of the quality and availability of data on those, but further research by HUJI suggests that they might not be demographically representative of other European cities, given their very high percentage of ethnic minorities.

Four options are proposed and discussed:

- Neukolln, Berlin, Germany
- Dinslaken, Germany
- Schilderswijk, The Hague, The Netherlands
- Kolenkit, Amsterdam West, The Netherlands
Prof. Weisburd asks ISTC-CNR what information is needed to move the modelling process forward. Dr Payette refers to the “Preliminary Terrorism Model Description” document (previously provided by ISTC-CNR), where the main questions are highlighted. In summary, these are:

- Which factors from WP 1-3 should be used as propensity and risk factors in the model, and how should they be weighted?
- What are the demographics characteristics of the population to be modelled?
- What types of beliefs would play a (positive or negative) role in the radicalisation process?
- What kind of locations (schools, mosques, etc.) should be included in the model and what kind of activities will the agents be performing at these locations?
- What are the specific policy questions to be addressed by the model?

Discussion moves to the question of the ‘threshold’ used to determine at which point an agent is considered to be radicalised for the purpose of model calibration and validation. The idea that this threshold can be calibrated empirically emerges from general discussion: using data about real-world radicalisation, we can set the threshold so that the number of radicalised agents produced by the model under normal conditions fits the empirical data. This should allow us to confidently use the same threshold for experiments involving alternative policy scenarios.

The partners agreed that ISTC-CNR would email a written list of questions to HUJI and HUJI would provide written answers to these questions.

The partners agreed that ISTC-CNR would email a written list of questions to HUJI and HUJI would provide written answers to these questions.

The partners agreed that ISTC–CNR would email a written list of questions to HUJI and HUJI would provide written answers to these questions.

The partners agreed that ISTC–CNR would email a written list of questions to HUJI and HUJI would provide written answers to these questions.

The partners agreed that HUJI would send written feedback to ISTC–CNR about the general content of the “Preliminary Terrorism Model Description” document.

The partners agreed that HUJI would send written feedback to ISTC–CNR about the general content of the “Preliminary Terrorism Model Description” document.

The partners agreed that HUJI would send written feedback to ISTC–CNR about the general content of the “Preliminary Terrorism Model Description” document.

The partners agreed that HUJI would send written feedback to ISTC–CNR about the general content of the “Preliminary Terrorism Model Description” document.

The partners agreed that the threshold for radicalisation in the model will be empirically calibrated.
Discussion moves to *policy questions* that can be implemented. Prof. Weisburd mentions the role of community centres as something that could be explored through simulation experiments. Prof. Calderoni suggests testing the (long-term) effects of returning foreign fighters. Prof. Hasisi mentions community policing, where police agents adopt strategies focused on creating stronger ties with community members. Dr Payette points out that this would require a more explicit description of the role of police agents in the model, which is currently unclear.

**Decision**

The partners agreed that the role of police agents in the model needs to be clarified further.

Still on the topic of possible experiments, Dr Andrighetto suggests that some general factors can be manipulated in the model: for example, the effect of having lower or higher values for collective relative deprivation, integration, employment, institutional trust, etc.

Some general discussion suggests that possible policy interventions could be framed in economic terms: given a certain budget, what is the optimal way of allocating resources in order to decrease radicalisation? Dr Payette points out that, in the interest of keeping the model simple, the cost of interventions could be calculated externally instead of being part of the computational model.

No final decision is reached regarding policy questions to be addressed by the model yet.

**Decision**

The partners agreed that HUJI will work out an explicit set of policy questions to be explored with the agent based model.

**Decision**

The partners agreed that HUJI and ISTC–CNR will start having virtual meeting once a month to work on the agent based model of recruitment to terrorism networks. The date of the first virtual meeting has been set to 2018-03-20.

The meeting ends at 17:00.
Section 2. Report on the operationalisation of factors into input for ABM simulations resulting from the activities conducted under task T4.2

Due to the decision of to develop separate simulations for organised crime and terrorist network Section 2 is divided into two parts:

- Report on the operationalisation of factors into input for ABM simulations on organised crime
- Report on the operationalisation of factors into inputs for ABM simulations on radicalization
Report on the operationalisation of factors into input for ABM simulations on organised crime

Author: UCSC - Transcrime

1. Introduction

This document reports on the activities conducted under Task T4.2 of WP4 on the operationalisation of factors identified during work packages WP1 to WP3 into rules for ABM simulations. Given the decision to develop separate simulations for organised crime (hereinafter OC) and terrorist network, this fragment focuses on the simulation of recruitment into OC.

Section 1 summarises the results of the activities conducted in WP1 and exposes the theoretical framework adopted for the development of the ABM. Section 2 presents the general structure and the main elements of the ABM on the recruitment into OC, Section 3 provides information on the data sources employed in the simulations, and Section 4 provides the ODD+D, a schematic document commonly used for ABM development to outline the main features of the simulations. Lastly, and since ABMs are employed for testing specific policies, Section 5 outlines the policy scenarios that will be tested through the ABM on OC.

2. From the factors for recruitment and criminalization into OC to input for ABM

2.1. Results of the systematic review

The systematic review of task T1.1 of WP1 synthesised the empirical evidence on the social, psychological, and economic factors relating to criminalisation and recruitment into organised crime groups (OCGs). Research findings are not only
relevant as the first systematic assessment of literature on recruitment into OC, but also because the resulting knowledge serves as input for the construction of the ABM.

The review searched and screened 48,731 potentially eligible records in five different languages, including the suggestions and contributions of scholars and experts. After a rigorous review process, the research team included 47 empirical studies employing quantitative, qualitative or mixed-methods approaches and relying on both primary and/or secondary data. Socio-economic and psychological factors associated with individuals’ recruitment into OCGs were organised into eleven categories:

- Age
- Gender
- Ethnicity
- Educational background
- Employment
- Economic conditions
- Social ties
- Group identity
- Psychological factors
- Criminal background and skills
- Silence/Omertà

The findings showed the prevalence of social and economic factors for recruitment into OCGs, while psychological factors were found to be marginal. Among the most commonly reported factors, the systematic review highlighted that individuals with violent attitudes and behaviour, low socioeconomic status, and kinship and blood ties with OC offenders are more likely to join OCGs. Further, such factors are highly interrelated, though their influence on individuals’ involvement into OC varies across types of OCGs. As for the geographic scope, most of the included studies investigated OC in the European context – namely Italy, the Netherlands, and Spain as the most represented...
countries – with less attention on other geographic areas also affected by OCGs’ presence (e.g. the Americas, Russia, Far East).

The systematic review revealed a rising trend in literature on recruitment into OCGs, with more than half of included studies published since 2010. The growth has involved all research methods, but most notably the use of quantitative methods. Despite the increased interest in factors associated with involvement into OC and the strong insights provided by the review, limitations of the literature exist. Limitations may be due, among others, to: the generalisability of the results beyond the OCG studied due to the heterogeneity of the concept of OC, the lack of external validity for case-study designs, and the issues related to non-random samples constructed from investigative/judicial data.

To overcome the limitations associated the literature included in the systematic review, the construction of the ABM of OC recruitment relied on the results from a broader literature and it also includes crime data from other sources. The information retrieved from the systematic review and a thorough review of literature on crime commission led to the formalisation and operationalisation of the rules that guide the agents through the model. The structure of the simulation will be described and explained in section 3.

2.2. THEORETICAL FRAMEWORK

The rationale of our simulated society is rooted not only in the outcomes of the systematic review on the factors leading to recruitment into OC (PROTON Deliverable D1.1.), but necessarily takes into account criminological theories and the related individual and social processes that lead to embeddedness in organised crime networks.

First, several theoretical perspectives suggest that organized crime is embedded in the social environment and that social relations are crucial for the recruitment into organized crime. For example differential association theory and social learning theory (see Sutherland 1937; Burgess and Akers 1966; Bruinsma 2014) posit that crime in its various aspects is learned in a social environment by relating with other criminal agents. This includes learning the material criminal
behaviours, the rationalisations surrounding and justifying them and, importantly, the internalisation of criminal identity aspects. Moreover, social and criminal embeddedness (Granovetter 1985a; McCarthy and Hagan 1995) suggest that the relations with criminal “tutors” mediating the learning process can be thought of as a form of capital for other agents. The position agents occupy within a criminal network determines their possibilities to commit crimes. In this sense, an agent’s valuable criminal ties determine his social opportunity structure (see Edward R. Kleemans and de Poot 2008a). This way, the number, strength and importance of an agent’s criminal ties strongly influences both his criminal embeddedness and output.

Second, other theories such as the general theory of crime (Gottfredson and Hirschi 1990), argues that an individual’s low self-control levels determine an inability to compute the negative consequences of one’s criminal behaviour, thereby determining persisting patterns of criminality throughout their life. The general theory of crime contends that group crime does not have specific characteristics and that the formation of criminal groups is mostly driven by self-selection processes.

While the social relations and self-control perspectives may generate opposing views about the recruitment into organized crime, the development of a agent-based simulation can easily include elements from both theoretical frameworks. In light of these considerations, the PROTON agent-based simulation for the recruitment into OC operationalizes criminal involvement both as the result of interaction with others and as emerging from agents’ natural propensity towards crime. The following section describes the technical formalisation of the framework within the simulations.

3. General structure of the OC ABM

Agent-based models are increasingly popular in the study criminal behaviour in general and criminal organizations in particular (Duxbury and Haynie 2019). Particularly, ABMs allow to test the impact of specific measures or policies that could hardly be experimented in the real world due to financial, ethical and social
constraints. This section presents the general structure of the ABM on the recruitment into OC, while the policies that will be tested in the simulation are presented in Section 4.

3.1. A MULTIPLEX NETWORK ABM

Simulating the dynamics and processes that lead to the recruitment into OC requires to take into account a wide variety of factors. As demonstrated by the systematic review conducted in WP1 of PROTON, recruitment into a criminal organisation often results from both individual and social factors. While certain factors are linked to an individual (e.g. age, gender), others inherently span over the personal characteristics of an agent: making new friends, for instance, is dependent upon the social environment in which an agent is set. Two children at the elementary school are more likely to become friends if they are in the same classroom or if they have the same age, rather than being separate into different classrooms or belong to different years. This is intrinsically related to the fact that the frequency of their contacts and interactions will be higher if they meet every day in the same classroom and they are thus part of the same little community.

In real life, every person engages in different types of relations, e.g. as part of a family, in friendships, at work, and in co-offending. Also, member of an organised criminal groups is part of a wider social environment, embedded in multiple social worlds. Relations of different types may drive the involvement and recruitment into organized crime (Arlacchi 1983; Arsovska 2015; Brancaccio 2017). This finding holds not only for Italian traditional mafias, where blood ties and kinship play a predominant role in explaining the influence of the social sphere in recruitment, but also for other types of groups, as gangs or Dutch organised crime (Decker and Chapman 2008; Edward R. Kleemans and De Poot 2008; Edward R. Kleemans and Van de Bunt 2008).

To adequately address the dynamics of individual and social drivers, we opted for an ABM based on multiplex social networks. A multiplex network includes several networks, each measuring specific social relations. The PROTON ABM for
the recruitment into organized crime includes the different types of connections considered relevant by the literature (as resulting from the systematic review conducted in WP1). Five relational layers are modelled in the simulations: families, friendship networks, professional and school ties, criminal relations and organised crime groups (Figure 1). Considering the need for creating a society that synthetically mirror the real-world, we have decided to include all the main dimensions through which an individual can realistically act and behave. The structure, topology and characteristics of the networks are empirically grounded using official statistics or replicating mechanics found in scientific works. The multiplex network framework also allows for considering individual-level characteristics as agents’ attributes, thus allowing to analyse individual and social factors to simulate realistic recruitment dynamics. Agents in the simulation, regardless of being or not part of an OC groups, can be born, get engaged/married, make children, die, create and break relations, and commit crimes.

Specifically focusing on the organised crime dimension, the model considers one single organised crime group existing in the simulation. Its internal structure, composition (in terms of gender distribution) and generation/age distribution reflects the ones found through analyses of several police investigations on OC groups.
3.2. RECRUITMENT INTO ORGANIZED CRIME

The main goal of the model is to simulate the recruitment into organized crime. For the purpose of this simulation, recruitment occurs when an agent commits a crime with at least another agent who is already a member of the OC group. This option was driven by different considerations. First, it is observable and easily operationalizable. Requiring the commission of a crime with OCG members models in a straightforward way the process of recruitment, avoiding subjective evaluations. Second, it is broadly consistent with the criminal law approaches criminalizing organized crime across EU Member States (Calderoni 2010). Most criminal justice systems require under their criminal law that, to be considered a member or participant in a criminal association, an individual participates in a crime committed within a criminal group.

Two different complementary dimensions contribute to the determination of the recruitment processes in the model, namely the probability of committing a crime (called “C”) and the embeddedness into organised crime (called “R”). In addition to determining the main outcome measure of the simulation (agents
recruited into the organised crime group), “C” and “R” provide additional outcome measures by yielding the overall number of crimes committed by all agents as well as all agents’ proximity to OC members (Figure 2).

![Diagram](Image)

*Figure 2. Synthetic description of the general structure of the model*

### 3.2.1. **The “C” Function**

The “C” function models the probability that agent $i$ will commit a crime at time $t$. Its structure revolves around two types of data source: the first source are official statistics and empirical information related to the specific location simulated by the ABM (see *Errore. L'origine riferimento non è stata trovata* section). The second source originates from the literature. Specifically, several systematic reviews providing information on the impact of certain factors on the general probability of committing a crime. These sources provide effect sizes (in different forms, e.g. odds ratios) allowing to determine the different probabilities of coming a crime given an agent’s network and individual characteristics. The list of individual factor-based rules that drive the model rules for committing a crime is presented in Table 1.
### Table 1. Individual level rules driving the crime commission process

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Rationale / Operationalisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age of the agent. Age is a well-known predictor of the risk of crime commission: we provide different risk based on age classes</td>
</tr>
<tr>
<td>Gender</td>
<td>Sex of the agent. Gender is one of the most comprehensively agreed risk factors for crime commission</td>
</tr>
<tr>
<td>Unemployment</td>
<td>Having/not having a job.</td>
</tr>
<tr>
<td>Education</td>
<td>Having/not having a high school diploma</td>
</tr>
<tr>
<td>Natural Propensity</td>
<td>Having a propensity score higher than a certain value (x) (need for assuming a statistical distribution since real data do not exist)</td>
</tr>
<tr>
<td>Criminal History</td>
<td>Having/not having committed a crime in the past</td>
</tr>
<tr>
<td>Criminal Family</td>
<td>Having a share of criminal family members (measured through ties, connections) which is higher or equal to 0.5. A criminal friendship/professional tie is a direct link with a family member which has committed at least one crime in the last 2 years.</td>
</tr>
<tr>
<td>Criminal Friends &amp; Co-Workers</td>
<td>Having a share of criminal friends (measured through ties, connections) which is higher or equal to 0.5. A criminal friendship/professional tie is a direct link with a family member which has committed at least one crime in the last 2 years.</td>
</tr>
<tr>
<td>OC Membership</td>
<td>Being already part of an OC group increases the probability of committing a crime by (\sim4) times.</td>
</tr>
</tbody>
</table>

Most crimes are committed by single offenders (Stolzenberg and D’Alessio 2008; van Mastrigt and Farrington 2009; Carrington and van Mastrigt 2013). Based on the literature on co-offending, a few crimes will require more than one offender. The literature on co-offending also helps in shaping the mechanisms that lead two or more individuals to commit a crime together. Peer and more in general social influence play a relevant role in driving this criminogenic cooperation (Weerman 2014). The model thus matches co-offenders based on mechanisms of social proximity: the closer two agents are in terms of social relations across network layers and the higher is the value of \(C\) of both individuals, the higher the probability of becoming co-offenders. The first crime committed by a given agent in association with someone belonging to an OCG is defined as the recruitment step of the given agent into the OCG.
C has been computed to provide realistic figures on committed offences within the model, in the form of rates by 100,000 inhabitants, using official statistics for different years (2012-2016, specifically). The calculation relied on the correction of crime figures by the dark number of each crime category. This allowed to take into account and include also those offences that have not been discovered or prosecuted in the original data at our disposal, thus giving more solid and reliable estimates.

Once agents commit crimes they can be incarcerated. Incarceration is estimated using empirical data retrieved from official statistics. Apart from family links, an agent in prison loses all ties that he/she has created during his/her life (including during his/her job). The mechanism for incarceration is based on a countdown that allows to establish when the agent leaves prison and returns to be “free” in the society, recovering part of its ties.

3.2.2. The Social Embeddedness into OC
R defines the embeddedness into OC to adequately model the probability of an artificial agent to be recruited into a criminal organisation. R seeks to consider the role of each agent’s social community in the process of recruitment. The theoretically-driven assumption is that individuals that are embedded in communities (across all types of networks considered by our simulation) that are highly populated by OC members face higher risks of being recruited. R should drive the selection of new OC members in the simulation. For example, among two equally suitable co-offenders, OC members are likely to co-offend with the agent who is more embedded in OC. In a relatively simple form - but coherently with the differential association and social learning theories as well as the social embeddedness of OC - R can be operationalised as the proportion of OC members among the social relations of each individual (comprising family, friendship, school, working and co-offending relations).

The proposed method implicitly weights the OC embeddedness as follows:
the importance of OC ties is inversely proportional to the distance. Thus having a OC member who is colleague of a friend is less important than having a colleague OC member.

- the importance of OC ties (but also of other non-OC ties) is proportional to the number of different ties between any two individuals. Thus, being a family member and a colleague of an OC member is more important than being just a colleague.

In addition to the contribution to determine recruitment into OC, the R function provides useful information to analyse the general dynamics of the model. First, R enables to clearly distinguish between active OC members and pro-OC agents. An agent may be strongly embedded in OC-prone networks but not necessarily be an OC member. For example, such agents could represent the role of women (e.g. wives and daughters of OC members) who are certainly living in OC-prone contexts but are rarely charged and convicted as OC members since they generally refrain from committing offences. Similarly, a juvenile son of an OC member who is just two years old is definitely not an OC member, but would still have a very high value of R, making it very likely that he will be recruited in the future. These agents would have a clear pro-OC role, “spreading” the social influence of OC. Second, R may contribute to the simulation of prevention policies, especially those on the primary and secondary socialization. The simulation will need to identify the target population and R could contribute in identifying the population at risk better than merely relying on other indicators e.g. the number of crimes committed by the parents or the involvement of a parent into OC.

4. Data

Using empirical data to feed the simulations is a fundamental step when aiming at setting up a reasonable and grounded model, besides theoretical and formal mechanisms (e.g. the mechanism of crime commission). Furthermore, data shall also be used ex-post to assess whether the model produces reliable and plausible results, especially considering the policy-oriented objectives of the ABM.
While the grant agreement required to create a single ABM for the recruitment into OC and terrorism, the consortium agreed that it would be difficult and hardly plausible to model the recruitment into both OC and terrorist network in a one simulation. The results of WP1-3 showed that the recruitment dynamics often differ significantly, suggesting to develop distinct AMBs for OC and terrorist networks. Furthermore, the recruitment into OC may be affected by the different individual and social conditions of a given society. For this reason, it was decided to run two different setups of the OC ABM, one modelling the city of Palermo (Italy) and one modelling the city of Leiden Eindhoven (the Netherlands).

To develop the model and validate the results, we had to retrieve and process several data from different sources regarding specific demographic, economic, social and criminal aspects. Since the City of Palermo is part of the PROTON consortium, the Palermo setup was developed as the first pilot. A synthetic table that includes the whole set of retrieved data, along with information related to the sources and the measured dimension is provided below.
Table 2. Data retrieve to develop and validate the ABM model

<table>
<thead>
<tr>
<th>Data</th>
<th>Type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data on the distribution of female fertility according to age</td>
<td>Demographic</td>
<td>Istat, 2017</td>
</tr>
<tr>
<td>Mortality probability by age and gender</td>
<td>Demographic</td>
<td>Istat, 2016</td>
</tr>
<tr>
<td>Wealth level distribution by a person’s level of education</td>
<td>Economic</td>
<td>Bank of Italy Survey, 2016 Istat, 2011</td>
</tr>
<tr>
<td>Data on employer sizes in city of Palermo</td>
<td>Economic</td>
<td>Istat 2011</td>
</tr>
<tr>
<td>Household size distribution by the household head’s age in Palermo</td>
<td>Demographic</td>
<td>Census data Istat, 2011 Municipality of Palermo</td>
</tr>
<tr>
<td>Distribution of household sizes in Palermo</td>
<td>Demographic</td>
<td>Census data Istat, 2011 Municipality of Palermo</td>
</tr>
<tr>
<td>Distribution of household type by age of household head</td>
<td>Demographic</td>
<td>Census data Istat, 2011 Municipality of Palermo</td>
</tr>
<tr>
<td>Distribution of people’s age by gender in Palermo</td>
<td>Demographic</td>
<td>Istat, 2018</td>
</tr>
<tr>
<td>Number of schools in the city of Palermo by level of education</td>
<td>Social</td>
<td>MIUR, 2016</td>
</tr>
<tr>
<td>Co-offending</td>
<td>Criminal</td>
<td>Istat, 2012-2016</td>
</tr>
<tr>
<td>Crime rates (corrected for dark number)</td>
<td>Criminal</td>
<td>Istat, 2012-2016</td>
</tr>
</tbody>
</table>

5. A draft ODD+D for the model of recruitment into Organised Crime

In summary, the model is structured as follows:

- Agents are in multiplex networks comprising family, friendship, school, professional, co-offending, and organized crime ties
- The simulations will rely on a theoretical framework that comprises differential association (Edwin Hardin Sutherland 1939) and social learning (Akers 2001). The framework is further integrated with the results found
in literature which show that individuals’ behavior – and also criminal and OC dynamics (in both intrinsic nature and recruitment) - heavily rely on social relations (Granovetter 1985b; McCarthy and Hagan 1995; E. R. Kleemans and Van de Bunt 1999).

- The intensity and multiplexity of network ties influence the probability of committing crimes and eventually get involved into organised crime
- Agents also have individual characteristics influencing tie formation processes and eventually criminal behaviors (e.g. gender, age, criminal propensity, wealth, education, ...).
- The recruitment into organised crime occurs when an individual commits a crime in co-offending with an agent who is already part of the OC group.
- Possible policies to be tested in the model:
  - Policies aiming at primary socialization (family): interventions on children in families at risk of organised crime
  - Policies aiming at secondary socialization (friends and work): education-based measures (opening schools to prevent the influence of deviant peers, promoting attainment of higher levels of education) and economic measures (measures to address income inequality, unemployment)
  - Policies aiming at the criminal socialization: different law enforcement disruption strategies (random or selective network disruption).

Following, we present our draft ODD+D of the model (Müller et al. 2013). An ODD+D is a protocol that expands the classic ODD (Overview, Design Concepts and Details) protocol adding a specific dimension on Human decision-making processes. In fact, the ODD was a protocol for describing and synthetizing the architecture of ABM originating from ecology, the ODD+D better suits model that involve individual and collective decisions made by human-like agents. Additionally, the ODD+D includes also a section on theoretical and empirical background in order to better frame the model itself. This document aims at
being a guide for the audience for easily understanding the main mechanisms and entities that are part of the simulations.
### Draft ODD+D Organised Crime

<table>
<thead>
<tr>
<th>Outline (→ template)</th>
<th>Guiding questions</th>
<th>What are the individual and relational factors that influence the processes of recruitment of OC members and development of OC groups?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overview</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I.i Purpose</td>
<td></td>
<td>The aim is to model the processes that lead individuals to the recruitment into OC based on their personal characteristics and social relations. The models will simulate two virtual societies, and two separate setups will mirror Palermo and Eindhoven.</td>
</tr>
<tr>
<td>I.i.a What is the purpose of the study?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I.i.b For whom is the model designed?</td>
<td>The model is designed for researchers and policy makers.</td>
<td></td>
</tr>
<tr>
<td>I.i Entities, state variables, and scales</td>
<td>I.ii.a What kinds of entities are in the model?</td>
<td>Entities include agents (that hold individual and relational attributes and characteristics), network layers (namely families, friendship networks, professional and school networks, criminal networks and an organised crime group) and also meta-entities such as firms.</td>
</tr>
</tbody>
</table>

---

Table 3. ODD+D Draft Table (The table uses the updated template for ODD+D from Müller et al. 2013)
I.ii.b By what attributes (i.e. state variables and parameters) are these entities characterized?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
|   | Individuals will be mostly characterized by two separate dimensions, $C$ and $R$, that will drive their actions and behaviors in the model. $C$ is the individual propensity to commit a crime that will be based on individual-level attributes. These attributes will include:  
- Age  
- Gender  
- Number of friends  
- Number of committed offenses / criminal history  
- Wealth level  
- Education score  
- Work status | Conversely, $R$ will map the relational dimension of each individual. Specifically, it will measure the extent to which an agent is embedded in OC-prone settings. For the sake of simplicity, $R$ can be seen as the proportion of OC members across the multi-plex $k$-step local networks. Edges of each network will also have specific characteristics regarding their respective weights. Indeed, the strength of an edge is inversely proportional to the distance.  

Note: the network simulation will rely on the concept of multiplexity. Indeed, different types of (weighted) links between agents (as friend or family ties, work and criminal relations) will mirror the complex heterogeneity of connections that exist within a given society. These connections are not independent from one another. For instance, family ties can influence the creation of a criminal ties in particular social settings (e.g.: areas where traditional mafias are established). A probabilistic function will model the risk of recruitment and OC involvement taking into consideration all the relations between contextual and personal variables (e.g.: types and strength of relations). |
| I.ii.c What are the exogenous factors / drivers of the model? | Exogenous factors take the form of law enforcement strategies or preventive policies. For example:

i) testing prevention policies in contexts at risk, e.g. policies aiming at reducing the negative influence on children in organised crime families (currently explored in Italy), keeping schools open also in the summer or incentivise pro-social behaviors from the early childhood (Does the intervention of educational institutions reduce the risk of being recruited?)

ii) testing law enforcement policies, e.g. different types of targeting strategies (does selective targeting rather than random targeting cause more damages to OC groups? Does disruption of OC groups cause fragmentation of violence? Does fragmentation increase violence?), |
<p>| I.ii.d If applicable, how is space included in the model? | Space is not included in the model. |
| I.ii.e What are the temporal and spatial resolutions and extents of the model? | Temporal resolution: each tick represents a given number of months. Note: Considering the data at our disposal and the processes that we intend to model, it would make little sense to adopt a shorter resolution (e.g.: hours or minutes). We intend to analyse long-term personal and social dynamics that are transferred from a generation to another, therefore the model needs a broader temporal resolution to be computationally feasible. |</p>
<table>
<thead>
<tr>
<th>I.iii Process overview and scheduling</th>
<th>I.iii.a What entity does what, and in what order?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agents act at every tick of the model with no determined order. Agents grow older, find jobs, go to school and establish or loose social relations of different types. Agents may commit crime. Law enforcement may arrest agents. Regarding the mechanisms of co-offending, a simple multi-step procedure has been proposed so far to model creation of new co-offending ties:</td>
<td></td>
</tr>
<tr>
<td>a. Identification of all nodes who commit a crime in a time period (based on empirical evidence of crime frequencies)</td>
<td></td>
</tr>
<tr>
<td>b. Identification of the share of all nodes who a) co-offend (based on empirical evidence of co-offending frequencies)</td>
<td></td>
</tr>
<tr>
<td>c. Based on b), identification of the number of co-offenders per crime (based on empirical evidence of co-offending size)</td>
<td></td>
</tr>
<tr>
<td>d. Matching of co-offenders based on social proximity</td>
<td></td>
</tr>
<tr>
<td>e. Establishment of new ties in the co-offending network</td>
<td></td>
</tr>
</tbody>
</table>
## Design Concepts

### II.i Theoretical and Empirical Background

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>II.i.a Which general concepts, theories or hypotheses are underling the model’s design at the system level or at the level(s) of the submodel(s) (apart from the decision model)? What is the link to complexity and the purpose of the model?</td>
<td>Two theories will constitute the backbone of the main framework of the model: differential association (Edwin Hardin Sutherland 1939) and social learning (Akers et al. 1979). Additionally, the recent stream of empirical research that proves the social embeddedness of organized crime further integrates the framework (E. R. Kleemans and Van de Bunt 1999; Edward R. Kleemans and de Poot 2008b). The theoretical setting posits that the propensity towards crime increases when an individual lives and acts within social contexts where there is an unbalanced relation between the acceptance of deviance and the rule of law. The more deviant is an individual’s community, the higher are the odds that this individual will start to commit a crime. Therefore, recruitment is not just a matter of individual propensity, but it is also driven by the interpersonal relations of each individual across multiple “communities” (k-step local networks). At the same time, individual attributes are relevant, in the sense that they may drive the risk of criminal behavior, as shown by scientific literature (Farrington, Gaffney, and Ttofi 2017). As an example, albeit an agent is particularly embedded in a pro-OC environment, it is highly likely that it will not commit a crime if its gender is female and its age is below 14. Conversely, a male agent that is not as OC-embedded as the former one but has specific individual risk factors (age, gender, economic status) might commit a crime and potentially become an OC-recruit candidate.</td>
</tr>
<tr>
<td>II.i.b On what assumptions is/are the agents’ decision model(s) based?</td>
<td>The core assumptions are related to both theories mentioned in the box above. Specifically, agents that are highly embedded in OC-prone environments are exposed to a higher risk of joining criminal groups. At the same time, findings from empirical literature demonstrates that certain individual characteristics heavily influence the propensity towards crime commission processes.</td>
</tr>
<tr>
<td>II.i.c Why is a/are certain decision model(s) chosen?</td>
<td>The focus on multiplex social relations aims at modelling the social and relational nature of OC and allow to control also for agent-level factors.</td>
</tr>
<tr>
<td>II.i.d If the model / a submodel (e.g. the decision model) is based on empirical data, where does the data come from?</td>
<td>Data come from different work packages (WP) of PROTON Project. Additionally, contextual data will be retrieved from other sources in order to simulate contextual factors that are not directly criminal. Specifically, data on SES conditions will be retrieved from national institutes of statistics and other related sources. Co-offending network properties will be calibrated based on the review of the quantitative empirical literature on the topic.</td>
</tr>
<tr>
<td>II.i.e</td>
<td>At which level of aggregation were the data available?</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>II.ii.a</td>
<td>What are the subjects and objects of decision-making? On which level of aggregation is decision-making modeled? Are multiple levels of decision making included?</td>
</tr>
<tr>
<td>II.ii.b</td>
<td>What is the basic rationality behind agents’ decision-making in the model? Do agents pursue an explicit objective or have other success criteria?</td>
</tr>
<tr>
<td>II.ii.c</td>
<td>How do agents make their decisions?</td>
</tr>
<tr>
<td>II.ii.d</td>
<td>Do the agents adapt their behavior to changing endogenous and exogenous state variables? And if yes, how?</td>
</tr>
<tr>
<td>II.ii.e</td>
<td>Do social norms or cultural values play a role in the decision-making process?</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>II.ii.f</td>
<td>Do spatial aspects play a role in the decision process?</td>
</tr>
<tr>
<td>II.ii.g</td>
<td>Do temporal aspects play a role in the decision process?</td>
</tr>
<tr>
<td>II.ii.h</td>
<td>To which extent and how is uncertainty included in the agents’ decision rules?</td>
</tr>
<tr>
<td>II.iii.a</td>
<td>Is individual learning included in the decision process? How do individuals change their decision rules over time as consequence of their experience?</td>
</tr>
<tr>
<td>II.iii.b</td>
<td>Is collective learning implemented in the model?</td>
</tr>
</tbody>
</table>
## II.iv Individual Sensing

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>II.iv.a What endogenous and exogenous state variables are individuals assumed to sense and consider in their decisions? Is the sensing process erroneous?</td>
<td>Individuals do not perceive any of their own characteristics.</td>
</tr>
<tr>
<td>II.iv.b What state variables of which other individuals can an individual perceive? Is the sensing process erroneous?</td>
<td>See line above. The sensing process can be erroneous and lead, for example, to conviction or death.</td>
</tr>
<tr>
<td>II.iv.c What is the spatial scale of sensing?</td>
<td>As pointed out in II.ii.f, space is not included, although network distances among nodes maps, in a certain sense, the “social” or community distance. Therefore, “spatial sensing” here can be thought as the threshold to which individuals (e.g. organized criminals) select nodes for certain actions. That distance is the $k$ that bounds the capacity to look for patterns or new recruits when a certain radius is exceeded.</td>
</tr>
<tr>
<td>II.iv.d Are the mechanisms by which agents obtain information modeled explicitly, or are individuals simply assumed to know these variables?</td>
<td>There is no information diffusion in the model.</td>
</tr>
<tr>
<td>II.iv.e Are costs for cognition and costs for gathering information included in the model?</td>
<td>No.</td>
</tr>
<tr>
<td>II.v Individual Prediction</td>
<td>II.v.a Which data uses the agent to predict future conditions?</td>
</tr>
<tr>
<td>----------------------------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>II.v.b What internal models are agents assumed to use to estimate future conditions or consequences of their decisions?</td>
</tr>
<tr>
<td></td>
<td>II.v.c Might agents be erroneous in the prediction process, and how is it implemented?</td>
</tr>
<tr>
<td>II.vi Interaction</td>
<td>II.vi.a Are interactions among agents and entities assumed as direct or indirect?</td>
</tr>
<tr>
<td></td>
<td>II.vi.b On what do the interactions depend?</td>
</tr>
<tr>
<td></td>
<td>II.vi.c If the interactions involve communication, how are such communications represented?</td>
</tr>
<tr>
<td></td>
<td>II.vi.d If a coordination network exists, how does it affect the agent behaviour? Is the structure of the network imposed or emergent?</td>
</tr>
<tr>
<td>II.vii Collectives</td>
<td>II.vii.a Do the individuals form or belong to aggregations that affect, and are affected by, the individuals? Are these aggregations imposed by the modeller or do they emerge during the simulation?</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>II.vii.b How are collectives represented?</td>
</tr>
<tr>
<td>II.viii Heterogeneity</td>
<td>II.viii.a Are the agents heterogeneous? If yes, which state variables and/or processes differ between the agents?</td>
</tr>
<tr>
<td></td>
<td>II.viii.b Are the agents heterogeneous in their decision-making? If yes, which decision models or decision objects differ between the agents?</td>
</tr>
<tr>
<td>II.ix Stochasticity</td>
<td>II.ix.a What processes (including initialization) are modeled by assuming they are random or partly random?</td>
</tr>
</tbody>
</table>
### II.x Observation

<table>
<thead>
<tr>
<th>II.x.a</th>
<th>What data are collected from the ABM for testing, understanding, and analyzing it, and how and when are they collected?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TBD:</td>
</tr>
<tr>
<td></td>
<td>Total number of families and kinship groups</td>
</tr>
<tr>
<td></td>
<td>Total number of friends</td>
</tr>
<tr>
<td></td>
<td>Total number of firms and work partners</td>
</tr>
<tr>
<td></td>
<td>Total number of offences, arrest and conviction rate</td>
</tr>
<tr>
<td></td>
<td>Prevalence of co-offending over the overall number of crimes</td>
</tr>
<tr>
<td></td>
<td>Estimated Total number of OC groups</td>
</tr>
<tr>
<td></td>
<td>Estimated Total number of convicted individuals</td>
</tr>
<tr>
<td></td>
<td>(...)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II.x.b</th>
<th>What key results, outputs or characteristics of the model are emerging from the individuals? (Emergence)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total number of citizens becoming members of organized crime groups (recruitment).</td>
</tr>
<tr>
<td></td>
<td>Number of offences</td>
</tr>
<tr>
<td></td>
<td>Co-offending prevalence</td>
</tr>
<tr>
<td></td>
<td>Total number of OC members</td>
</tr>
<tr>
<td></td>
<td>Total number of convicted individuals</td>
</tr>
<tr>
<td></td>
<td>OC-offending prevalence (out of the overall number of crimes)</td>
</tr>
<tr>
<td></td>
<td>Dynamic network metrics in order to evaluate the evolution of the network after a given policy is implemented</td>
</tr>
<tr>
<td></td>
<td>Note: Emergence is modelled through different dimensions at individual (micro), group (meso) and environment level (macro). All these dimensions are interconnected. At individual level, emergence regards the degree of frequency of a criminal career or the role acquired or the possibility to freely act in the system (which is limited by two states: conviction and death). At group level, emergence is related to the intra-equilibrium within a single mafia network (e.g. after a policy is implemented)</td>
</tr>
</tbody>
</table>

### III Implementation Details

<table>
<thead>
<tr>
<th>III.i</th>
<th>How has the model been implemented?</th>
</tr>
</thead>
<tbody>
<tr>
<td>III.i.a</td>
<td>It will be implemented in NetLogo (using also the R extension RNetLogo).</td>
</tr>
</tbody>
</table>
## III.i.b Is the model accessible and if so where?

The open access code of the model will be accessible on the PROTON website. Main results of several iterations and model specifications will be shared through a Wizard and will aim at giving practitioners and policy makers a tool for understanding how recruitment (and related criminal dynamics, e.g. number of committed crimes) changes after certain policies or certain structural situations are tested.

### III.ii Initialization

| III.ii.a What is the initial state of the model world, i.e. at time \( t = 0 \) of a simulation run? | TBD. Note: Initialization is allowed to vary for sensitivity analysis processes. Parameters will be changed along with random seeds. |
| III.ii.b Is initialization always the same, or is it allowed to vary among simulations? | The initial values will be chosen based on data. It might be also interesting to use arbitrary/random data to test whether the model produces effects that are based on the conditions at the starting point or not. |
| III.ii.c Are the initial values chosen arbitrarily or based on data? | External data used in the model are provided by the partners of PROTON Project and are then processed to result in the formalisation of the model. |

### III.iii Input Data

| III.iii.a Does the model use input from external sources such as data files or other models to represent processes that change over time? | External data (also to model potential variations and processes over-time, for instance in the case of SES conditions) will be retrieved from external open access sources in most cases. For a detailed list, see section *Errore. L'origine riferimento non è stata trovata.* |

### III.iv Submodels

| III.iv.a What, in detail, are the submodels that represent the processes listed in 'Process overview and scheduling'? | / |
| III.iv.b What are the model parameters, their dimensions and reference values? | / |
|   | III.iv.c How were submodels designed or chosen, and how were they parameterized and then tested? | / |
6. Policy Scenarios
The final goal of PROTON-S simulations is to test potential policy scenarios that may help enhance the knowledge on how to reduce and counter the involvement of individuals, especially youth, into criminal organisations. The decision of the scenarios to be tested followed the suggestions and remarks given by the consortium partners during consortium meetings and workshops as well as ad-hoc meetings organized with policy makers held in September 2018 in Milan and Amsterdam. It was decided to test two different types of policies, namely:

1. Primary and secondary socialisation (family and school)
2. Law enforcement

6.1. PRIMARY AND SECONDARY SOCIALISATION SCENARIO
This policy scenario refers to a number of prevention policies addressing school-age agents in the model to keep them from criminal activities and organized crime recruitment. In particular, the central policy aims at protecting juveniles from socialization processes leading them to recruitment into OC. Among the possible solutions, policies aiming at decreasing or countering the influence of OC-prone social relations may offer an effective strategy to prevent the recruitment into OC. Such policies may include interventions including welfare measures, psychological support, and addressing the parents’ negative influences on juveniles.

In the simulation, these policies will be modelled through different solutions. They may include reducing the influence of parents (normally the father), addition of pro-social ties such as non-criminal friends, provision of educational and job opportunities (increasing education achievement, wealth and work ties). This scenario will put to the test and compare the effectiveness of parental authority suspension and other support policies toward reducing crime rates and OC recruitment, both as separate and combined treatments.

Policy scenario:
The simulation will model policy scenarios related to primary socialization and secondary socialization targeting individuals living in contexts of OC.

The primary socialization policy aims at young people aged 12-18 living in OC families, intended as families where at least one parent is an OC member. It will be possible to select the share of the target young people based on a risk score reflecting the family embeddedness (R) in OC and/or based on the OC members convicted in the simulation. This group of juveniles at risk will then be subjected to interventions measures aimed at reducing OC parental ties. For instance, the simulation will be able to model cases in which court orders limit the contacts between people involved in OC and their families or cases of OC members’ conviction and imprisonment. In such cases, the ABM will temporarily decrease the relation that OC members have with their families and children while also providing juveniles at risk and their mothers with social and welfare support (e.g. school support, employment support).

The secondary socialization aims at children and young people aged 6-18 who are in school. Crime-prone children, i.e. those with higher criminal propensity (c score), will be targeted with increased social support and/or increased welfare support. Increased social support may include: 1) better educational support (in the ABM, the agent will complete high school and/or achieve a higher level of education), 2) support of psychologists and social workers (promotion of prosocial relations and inhibition of anti-social relations, randomly creating friendship ties with non-deviant peers and adults), 3) increased social activities between children (children will randomly create new friendship ties), and/or 4) move child to new class (if school classes are present in the ABM, this should generate additional friendship ties). Increased welfare support instead may include providing a job to the child’s mother (resulting in diversification of mother’s networks) and/or providing the child with a job when they turn 16 (resulting in diversification of the child’s social networks and lower risk of crime commission).
6.2. Law Enforcement Scenario

This scenario aims at analysing the impact on the recruitment into OC of different law enforcement strategies in tackling OCGs. In particular, the possible targets will consist of OC group bosses/lieutenants (in the model, potentially OC agents with high scores in measures such as betweenness) and workers in “facilitator” positions. “Facilitators” include logistic workers, such as long-distance truck drivers and airport workers, and legal and financial advisors. These agents have increased opportunities for crime thanks to their work position (e.g. drug smuggling, money laundering). This scenario will put to the test and compare the effectiveness of LEAs targeting OC bosses/leaders and facilitators toward reducing crime rates and OC recruitment, both as separate and combined policies.

Targets:

1. Workers in facilitator positions
   - Policies:
     - Higher scrutiny of facilitator positions (will decrease facilitators’ probability of crime commission and OC tie creation)
     - Higher repression of criminal facilitators (will lead to higher imprisonment rates for criminal facilitators)

2. OC bosses/leaders:
   - Policies:
     - Higher scrutiny of OC members (will decrease their ability to commit crimes and consequently create OC ties)
     - Higher repression of OC members (will lead to higher imprisonment rates for OC members).
7. References


Report on the operationalisation of factors into inputs for ABM simulations on radicalization

Author: HUJI

1. Introduction Terrorist Network

This section of T4.2 reports on the development and operationalization of the Agent Based Model (ABM) foreseen in WP5. Specifically, T4.2 reports on how the results from WP2 have contributed to the development of the ABM for the radicalization-recruitment to terrorism model. WP2 was led by a systematic review and meta-analysis (T2.1) which included the studies T2.5 and T2.7. The review also included an extensive literature review of the risk factors for radicalization and recruitment to terrorism based on studies which were not included in the meta-analysis. The results of the review have served as the basis for the development of the ABM's theoretical framework and mechanisms, while the results of the meta-analysis have provided the statistical basis for the ABM's inputs.

As foreseen in the project PROTON structure and plan, WP5 consists of an ABM that will be used to examine different experimental scenarios which cannot be experimented with in real-world settings. As an exploratory simulation study, the ABM will be used to conduct a select set of experiments—as discussed in T4.1—to test the effects of three key policy changes in a simulated world in which a number of select risk factors are in operation; targeted employment, increased numbers of community workers, and community policing. Each of these policies have both direct and indirect effects on the model's mechanisms and risk factors.
2. Results of the systematic review

T2.1 included the synthesis of data pertaining to 61 risk and protective factors across 3 outcomes. While the majority of included studies were from Europe and focussed on Islamist-inspired cognitive radicalization, moderator analyses found little differences between effect sizes from non-European estimates and other radical doctrines such as right-wing extremism. The results of T2.1 highlight the difficulty in classifying risk factors according to broad categories such as economic, social and psychological. Most factors represent hybrid categories, such as socio-economic, or social-psychological. However, it is possible to say that factors are generally either individual characteristics, experiential, environmental, or factors related to subjective opinions on different topics.

Given that ABMs require a high degree of parsimony, it was not possible to include all 61 factors in the model (T4.1). Additionally, as many factors are not subject to change, modelling their effects over time would not provide any new insights into the way in which they affect radicalization. As such, a theoretical model guided the selection of the specific factors that would serve as the model's inputs. The theoretical framework was informed by best practices for ABMs in criminology, the literature review of T2.1, as well as the empirical results of the other studies from WP2.

3. Theoretical Framework

The theoretical framework developed from the results in WP2 is based on an integration between the attitude-behavior continuity model (Ajzen & Fishbein, 1980), which also serves as a basis for the two-pyramid of radicalization model (Moskalenko & McCauley, 2009; McCauley & Moskalenko, 2017), and a risk-
protection factor paradigm (Stouthamer-Loeber et al., 1993; Hawkins et al., 2000). The model suggests that there are two dimensions to radicalization: attitudes and behaviors. According to the attitude-behavior continuity, attitudes are one of the greatest predictors of future behaviors. The results of T2.1 provide statistical evidence that supports this model with respect to radicalization. However, only a small minority of those with radical attitudes will ever engage in radical behaviors. According to a risk factor-protective factor framework, the development of radical attitudes, and the move from radical attitudes to radical behaviors, is the result of the cumulative weight and interactive effect of multiple risk factors over the existence of, or in absence of the buffering effects of protective factors (T2.4). As identified in T2.1, factors generally fall into the categories of characteristics, experiential, psychological, or opinion based factors. While most individual level characteristics, experiential, and psychological factors are either not open to change or are otherwise unlikely to be affected by policy, opinion based factors may be.

As per the results of T2.1's literature review, and drawing on the most popular theories of radicalization, the development of radical attitudes occurs primarily through socialization, in which individuals are exposed to different radicalizing messages from peers, family, members of the community, and the internet (Hegghammer, 2013; Holt, 2012; Weimann, 2011; Hamm & Van de Voorde, 2005). So too, the development of, and changes in different opinion based risk and protective factors, such as legitimacy, is primarily a function of socialization and differential associations (Walters, 2016a,b). As demonstrated by the results of T2.1, differential associations (peers) are in and of themselves an important risk factor both for radical attitudes and radical behaviors.

However, differential associations have a high correlation with environment and social structure (Akers, 1998). Individuals engage in routine activities, such as going to work, school, public places, and commuting. These routines generally follow a certain pattern based on an individual's demographics and
characteristics. These routine activities condition, limit, and determine differential associations, and differential associations also play a role in determining routines (de Waele & Pauwels, 2014; Pauwels and Svensson, 2013). As such, changes in an agent's routine activities may stem from changes to either their individual characteristics, or changes to their environment. For example, the employment of a previously unemployed individual will lead to significant changes in their routine activities. Changes in routine activities will necessarily impact and change the availability of differential associations, and the opportunities for interactions with them. As such, in the example of an unemployed individual gaining employment, this will lead to "new opportunities for socialization and changed routine activities" (Laub & Sampson, 2009:43).

This differential availability of differential associations, who provide differing degrees of support for normative or deviant attitudes and behaviors, is dependent on environmental conditions (Cloward & Olin, 1960; Akers, 1998). Individuals learn attitudes (and behaviors) from differential associations who are determined by their routine activities within their social structure and environment. Once learned, these attitudes (and behaviors) are also incorporated into routine activities in a process summarized by Hindelang et al. (1978:242):

Structural constraints originating from [the social structure] can be defined as limitations on behavioral options that result from the particular arrangements existing within various institutional orders, such as the economic, familial, educational, and legal orders. For example, economic factors impose stringent limitations on the range of choices that individuals have with respect to such fundamentals as area of residence, nature of leisure activities, mode of transportation, and access to educational opportunities.
For a susceptible individual to come into contact with radicalizing influences, however, they must be exposed somehow; by happenstance, by way of their social networks, or through an active seeking process. Interactions with radicalizing influences, regardless of how contact occurs, occur within built environments and locations; such as a mosque, pub, community centre, or a virtual environment, such as the internet (Bouhana & Wikström 2011). Its these interactions that determine radicalization (attitudes) and recruitment (behavior).

4. General structure of the ABM

The model represents a city borough comprised of four different communities with different populations. In addition to citizens, the model includes different locations, such as private residences, open spaces, places of employment, religious institutions, and community centers. Citizen agents move from one location to another to perform different activities based on their individual level characteristics. Some of these activities are mandatory, such as employed people spending a specified portion of time at work, and another amount of time sleeping and being inactive. Other activities are optional, such as commuting from one location to another, leisure activities, attending a religious institution, going to activities at a community center, socializing, and consuming media.

As part of these activities, citizens engage in communications with other agents, either passively, actively, or both. When agents communicate, they discuss certain topics which are the objects of the interaction. For example, they could speak about a general topic, such as “I disagree with the government’s new policy” which may influence the opinion of the other. Or they could discuss a particular location, saying something like “You should come to the community center”, which may influence the other's decisions to attend such a location. They could also discuss another individual, such as “Alice is a good person”, which may influence the likelihood that the other will engage in future interaction
with Alice. Through these communications, agents can influence each other’s opinions and decisions.

Not all interactions involve the possibility of the influence going in either direction. For example, when an agent consumes media, the interaction is purely passive and only the agent has the potential to be influenced. Similarly, certain agents, such as religious leaders, community workers, and police, serve specific functions in which they are actively trying to influence the attitudes and behaviors of citizen agents. They themselves do not change their own attitudes and behaviors based on their interactions with ordinary citizens.

These elements of the model will be implemented in NetLogo, a software designed for ABMs. NetLogo models involve three types of agents: turtles, patches, and links. Turtles are called “turtles” for historical reasons (Tisue and Wilensky, 2004), but they can simply be thought of as 'mobile agents' who are able to move around. NetLogo provides a space that is a 2D plane supporting floating point cartesian coordinates. Patches are square cells that are superimposed on that plane. While patches remain static, as they represent geography, they can be characterized by different features and variables. Links represent the relationships that exist between turtles. Links can also have specific types of behaviors and be characterized by different variables. Turtles and links (but not patches) can have different “breeds”. Each breed can have different variables and behaviors. Furthermore, in the case of links, breeds can be “directed” or “undirected”. Links of an undirected breed are always reciprocal. For example, if agent A1 has an undirected link with agent A2, then A2 shares that undirected link with A1. Conversely, directed links are unidirectional. For example, if A1 has a directed link to A2, while possible, it is unnecessary that A2 have a directed link to A1.

When we say that an agent “has” or “owns” a variable, it means that each agent has a separate value for that variable. For example, agents of a specific age group, immigrant status, or employment status will have individual values for
each of these variables. NetLogo models operate an internal clock that “ticks”. Each “tick” is a step in the execution of the model’s main algorithm. At each tick, the values of each variable for each agent is updated.

4.1. COMMUNITIES
The model represents an area of a European city, a borough, with four distinct and neighboring communities. Some global parameters of the model are defined regarding the number, size and population of the communities. While the size of a patch in meters is not formally defined, it is taken to be the size of one residence. The number of citizens per community is defined by the citizens-per-community parameter and citizens are randomly assigned to residences within the community. While this represents an abstraction of a city—since real cities are not square and communities vary in shape in size—the assumption is that the geographical features captured by this abstraction (relative position of communities to each other, population densities, distances between locations etc.) are those relevant to the processes being modelled.

Table 4: Global parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Domain</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>num-communities</td>
<td>{ n^2</td>
<td>(n \in N_0) \land (n^2 \mod 2 \neq 0) }</td>
</tr>
<tr>
<td>community-side-length</td>
<td>N_0</td>
<td>The side length of a community in patches.</td>
</tr>
<tr>
<td>citizens-per-community</td>
<td>N_0</td>
<td>The number of citizens in each community.</td>
</tr>
<tr>
<td>activity-radius</td>
<td>N_0</td>
<td>How far (in patches) agents look around them for free activities. See section 2.4</td>
</tr>
<tr>
<td>alpha</td>
<td>[0,1]</td>
<td>How strongly tolerance and strength of opinion are related. See section 3.2.</td>
</tr>
<tr>
<td>radicalization-threshold</td>
<td>[0,1]</td>
<td>The sum of risk factor values necessary for a citizen to become radicalized.</td>
</tr>
</tbody>
</table>

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 699824.
Citizens can and do overlap at particular places but since they are represented as slightly transparent, the opacity of the figure at a particular tick indicates the population density in a given patch (Figure 1). While theoretically there are no limits as to how many agents can exist in the model, computational requirements and demand grow nonlinearly, and processing requirements and time need to be taken into consideration. As such, the model includes a downsized population of 20,000 citizen agents (T4.1).

Figure 4: Representation of agents in community

4.2. SETTING UP LOCATIONS
The particular locations that exist in each community are defined as part of a model scenario. The number and types of locations vary from one community to the next based on official data sources pertaining to the chosen borough that serves as the basis for the model. Each community is assigned a set number of residences, open spaces, places of employment (including number of employees), community centers, mosques, and bars. While the number of the different location types are based on official data, their locations are distributed in the community according to a weighted random selection of patches.

4.3. SETTING UP CITIZENS AND THEIR ATTRIBUTES
There are five different types of agents in the model; ordinary citizens, community workers, religious leaders, police, and already recruited agents. The number of citizens of each category that exist in each community are determined according to official data sources used for creating the initialization distributions. Each citizen is assigned a residence at random, with multiple citizens being assigned to the same residence to represent households. With regards to ordinary citizens, demographic data is used to create the distributions of the agents’ characteristics such as age, gender, religion and immigrant status. The data is matched with survey data in order to calculate the distributions and assignment to agents of different scores for each of the risk factors for both the propensity and risk formulas.

While propensity factors and risk factors, and their formulas share the same goal of computing an individual’s “risk of radicalization”, they differ in that propensity is a fixed score, whereas risk is dynamic. The factors that are included in the propensity formula represent static or historical factors (such as age and immigrant status), whereas the factors included in the risk formula are dynamic, opinion based factors. Each of the factors carry different weights, based on the results of T2.1. Once the risk of radicalization reaches a threshold defined by the radicalization-threshold parameter, the agents can be classified as having been radicalized. When radicalized agents reach a certain threshold of relationship with an already recruited agent, they can be said to have been recruited. Official data on the number of recruited individuals that exist in the
modelled borough is used to install pre-recruited individuals in the model at initialization.

Community workers, religious leaders, police, and already recruited agents are assigned fixed scores and they have no propensity or risk. Community workers and police have wholly positive scores on each of the opinion based factors, whereas already recruited agents have wholly negative ones. Religious leaders are split in their opinion based factors' scores as per official data sources identifying the split between extremist and moderate institutions that exist in the real-world borough serving as the basis for the model.

4.4. **Activities**

The model simulates the daily lives of citizens living in communities. Citizens create "activity links" based on their routine activities and past activities and a value is assigned to each activity for each agent. The model includes three different types of activities: free time activities, mandatory activities, and jobs. Free time activities can be performed at any time a citizen is not doing a mandatory activity, or at their job. These activities implicitly have a duration of one hour, but can be performed again in consecutive hours if the citizen wishes to do so. At the beginning of a simulation, citizens only know about free time activities that are located around the locations of their mandatory activities: in practice, that means near their residence or work. As the model progresses over time, agents can learn about other possible activities through discussions with other citizens.
Mandatory activities have to happen at a particular time of day and have a specific duration in hours. Thus, for example, everyone has to go home by midnight and remain at home for at least eight hours. In mandatory activities, the activity has to be performed at the location of that type that is the closest to the citizen’s residence.

Citizen agents' routines are characterized by activities, most of which involve communication with other agents. While communication usually occurs between two agents, certain agents are able to communicate with multiple citizen agents at once (e.g. Community workers, religious leaders and police). The process is the same in both cases however: one agent talks about something, and the other(s) listen. Their opinion-related factors' scores may change as a result of the interaction. To simulate that process of opinion transmission, a variant of the Gargiulo and Mazzoni (2008) model is used. In the model, the value of an opinion can range from −1 to 1, where −1 expresses disagreeing or disliking and 1 expresses agreeing or liking. Opinions held by citizens on various topics can constitute risk or protective factors depending on the direction and strength of the value and as per the weight for each factor as derived from T2.1. Citizens receive a score for each topic at initialization, which are assigned to agents based on distributions and correlations with other characteristics as derived from official survey data.
In the model, interactions are directed: agent \( a_i \) tells agent \( a_j \) about something, and \( a_j \) can update its opinion or not. Whether or not an update occurs depends on \( t_j \), and the tolerance of \( a_j \). That tolerance is a function of the strength of \( o_j \). The opinion of \( a_j \), is taken from the value of a topic or activity link. Opinions near the extremes (−1 and 1) are considered stronger than neutral opinions (near 0). The effect of the strength of \( o_j \) is mediated by \( \alpha \) (the alpha parameter). The interaction will be successful of the initial difference in opinion between the agents is less than the tolerance. If the interaction is successful, \( o_j \) is updated “towards” \( o_i \), by an amount that also depends on \( t_j \). The same mechanism is used for “broadcast” activities like preaching or teaching. The only difference is that instead of only talking to a single agent, agents who can broadcast talk to a set of receiver agents and all receivers take the role of \( a_j \) in parallel.

Figure 8: Opinion dynamics

5. Inputs
In order to incorporate as many factors into the model as possible, whilst maintaining parsimony and computational optimality, a limited number of factors were selected from each category. Each factor was chosen for its theoretical relevance to the model and its position as an important risk factor as per the findings of WP2. Factors were also chosen based on the availability of external data to calculate their distributions within the population and to provide for validation.

We sought to incorporate the most important opinion based factors. In the results of T2.1, only two factors had independent protective and risk effects; integration and law legitimacy/trust. As a result of having individual protective
and risk effects, a -2 to +2 scale for these variables was created. While T2.1 indicated that objective economic factors were not significant, relative deprivation was. While there was no protective factor associated with relative deprivation (and no theoretical protective factor exists either), it was important that the model include an element relating to the economic dimension. As such, relative deprivation—*subjective deprivation*—was chosen as the third factor and placed on a 0 to +2 scale.

For individual characteristics, we selected age, gender, and immigrant status as these factors are readily available in external data sources and are relevant to all population. For experiential factors we selected criminal history as WP2 found this factor to be of particular importance and data pertaining to the distribution of criminal history in the population was also available. With regards to psychological factors, the results of T2.1 indicate that some 'innate' characteristics are among the most important risk factors. For example, low self-control and authoritarianism/fundamentalism.

Both low self-control and authoritarianism have been conceived to be innate traits (Adorno et-al, 1950). While available data would make it difficult to identify the distributions of levels of self-control in a population, survey data makes the development of distributions for authoritarianism/fundamentalism possible. It is important to note that while more recent discourse has come to accept that authoritarianism is learned through socialization, it remains that there is good evidence to suggest that some are more prone to it than others. Given the availability of data, and the computational considerations, we selected authoritarianism/fundamentalism as the psychological propensity factor to be included in the model.

**Table 5: Factors in the model**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, Employment, Immigrant status</td>
<td>Individual characteristic</td>
</tr>
</tbody>
</table>
6. Data

Unit of analysis: City borough

As discussed in T4.1, it is quite difficult to determine what constitutes a "prototypical" city. Major European cities differ from each other significantly in almost every category. However, city boroughs tend to share a lot of common characteristics, in part due to the fact that boroughs are more homogenous than cities are (albeit still quite heterogeneous). Additionally, evidence suggests that local communities are the primary places where radicalization and recruitment occur. Trends in terrorism have highlighted the fact that certain boroughs tend to be responsible for producing multiple terrorists, and radicalization tends to be concentrated in certain boroughs over others. It is also understood that boroughs are the primary geographic level at which contextual factors influence and moderate individual factors vis-à-vis radicalization and efforts to stymie it (Slootman & Tillie, 2006). These understandings are reflected in the local, community based approaches currently being undertaken by European countries in the framework of countering violent extremism (CVE) initiatives and policies.

The importance of smaller simulation geographies such as the borough as a unit of analysis is well reflected in the use of ABMs in criminology (Weisburd, Braga, Groff, Wooditch, 2017). Focussing on boroughs limits the effects of confounding variables and helps address issues of computational complexity (Weisburd, Wooditch, Weisburd and Yang, 2016). At the same time, the current simulation is of a representative landscape, with representative characteristics. So simply scaling it up by including more geographic units should produce outcomes that are still representative to larger geographies, and other cities. In this regard, Bosse, Elffers, and Gerritsen (2010) examined directly whether small-scale ABM models produce results similar to larger simulations. The authors doubled the
size of their model and found, “the effect of scaling up the size of the society turned out to be small” (Bosse, Elffers, and Gerritsen, 2010: 63).

6.1. SETTING: NEUKÖLLN

Neukölln is one of the largest of Berlin's twelve boroughs, but it is also one of the poorest and has the highest proportions of immigrants. Neukölln includes many of the features relevant to the current study. Firstly, Neukölln is split into four distinct areas, each of which are considerably different from each other. It was found that rich city level and borough level data was available from which we could model the characteristics and populations for each of these areas. Secondly, while Neukölln is widely seen as a potentially high-risk borough (Shoshan, 2008; Soederberg, 2017) and has produced some extremism (both religious and right-wing inspired), the numbers have still been relatively low. It therefore provides a good representation of the concerns of policy makers; that high-immigrant and low-income areas which have yet to produce much radicalization and recruitment may do so in the future.

Thirdly, Neukölln’s community based approaches (which have yet to be evaluated) to countering radicalization and violent are representative of similar efforts being taken in other major European cities and their boroughs (Berczyk, 2013). These community-based policies include a 2010 initiative to improve integration through the provision of public services as well as employment (Council of Europe, 2018). Additionally, while Neukölln has both religious institutions known for involvement with extremism, others have long cooperated with official bodies in the fight against extremism (OSCE, 2014). However, it remains unknown if these programs have any positive effect (Buschkowsky, 2013). As such, Neukölln can serve as a representative area to be modelled which can provide results that are highly generalizable to other European cities.
6.2. DATA

Data was derived from multiple sources for the different components of the model. For the population and landscape characteristics, data was derived from a number of datasets provided by the Berlin bureau of statistics (Statistik Berlin-Brandenburg), namely the 2011 Berlin Census and Microcensus datasets. Data on the number and type of politically motivated events and arrests were obtained from "Lagedarstellung Politisch motivierte Kriminalität in Berlin 2017" report produced by the Berlin Police. Data on the number and type of encounters with police in Neukölln were derived from the 2017 Neukölln Criminality Report.

In order to create the initial distributions of the opinion topics, we are currently working with three different datasets in order to identify which one provides the most suitable representation. These datasets are: The European Values Survey (EVS) for Germany (2008 and 2017), The German General Social Survey, 2008 (ALLBUS), and the Social Relations and Conflict Potential in the Context of Experiences of Denied Participation and Appreciation of Youth with and without a Migrant Background, 2015 survey (ZA5169).

Table 6: Data sources for the different factors in the model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data source</th>
<th>Effect source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>2011 Berlin Census</td>
<td>T2.1</td>
</tr>
<tr>
<td>Gender</td>
<td>2011 Berlin Census</td>
<td>T2.1</td>
</tr>
<tr>
<td>Immigrant status</td>
<td>2011 Berlin Census</td>
<td>T2.1</td>
</tr>
<tr>
<td>Religion</td>
<td>2011 Berlin Census</td>
<td>T2.1</td>
</tr>
<tr>
<td>Employment status</td>
<td>2011 Berlin Micro-census</td>
<td>T2.1</td>
</tr>
<tr>
<td>Criminal history</td>
<td>2017 Neukölln Criminality Report</td>
<td>T2.1</td>
</tr>
<tr>
<td>Contact with police</td>
<td>2017 Neukölln Criminality Report</td>
<td>T2.1</td>
</tr>
<tr>
<td>Authoritarianism/ Fundamentalism</td>
<td>EVS/ALLBUS/ ZA5169</td>
<td>T2.1</td>
</tr>
<tr>
<td>Integration</td>
<td>EVS/ALLBUS/ ZA5169</td>
<td>T2.1</td>
</tr>
<tr>
<td>Legitimacy/Trust</td>
<td>EVS/ALLBUS/ ZA5169</td>
<td>T2.1</td>
</tr>
<tr>
<td>Subjective deprivation</td>
<td>EVS/ALLBUS/ ZA5169</td>
<td>T2.1</td>
</tr>
<tr>
<td>Number of recruited terrorists</td>
<td>Berlin Police Report on Politically Motivated Violence; Background to the relatives of the Salafistic spectrum in Berlin, 2017</td>
<td>N/A</td>
</tr>
</tbody>
</table>
 Associations with recruited terrorists | As will occur in the model | T2.1

*More recent versions of these reports and surveys will be used for validation purposes.

### 7. Policy scenarios

#### 7.1. Employment

One element of the integration policies in Neukolln has been employment. However, it is currently unknown what the effect of this is on radicalization. The results of T2.1 indicate that while un-employment is a relatively small effect risk factor for radicalization, it may be more important as a risk factor for recruitment. Additionally, employment status has a significant effect in determining and shaping routine activities, and thereby differential associations.

Using the data from the Berlin City for Neukolln, we have a distribution of employed and unemployed individuals. We will select a percentage of currently unemployed individuals from the base model who have high risk of radicalization scores and assign them to employment at the initialization of the experimental model.

In this experiment a government-led initiative would match employers with a targeted population of unemployed individuals who are considered to be high-risk. The employers would be offered incentives to hire such individuals. Individuals will be considered high-risk based on a risk assessment that will look at factors such as: Age (16-30), immigrant status (first generation), education (high-school graduates).

In order to run this experiment, the selected individuals will become employed at the simulation's initialization.

The primary effects of this model will be:
Direct-On the propensity equation

Indirect- On the routine activities of the agent which determine differential associations.

This simulation can be run identically for both Islamic and Right-Wing groups, with the exception of the immigrant status variable's inclusion in the risk assessment.

The experiment would compare two scenarios:

1) World without the employment program.

2) World in which all individuals who meet the threshold are placed in the program.

7.2. COMMUNITY CENTER WORKERS

In this experiment, we attempt to reduce radicalization and recruitment by increasing the number of community workers. Community workers are assumed to encourage values that lower risk of recruitment. We do not construct new community centers which is likely unrealistic as a policy approach.

The weight of the influence of community center workers will need to be determined but they will have fully positive scores across all of the opinions operating in the opinion dynamics functions.

The positive influences of the community center workers will affect those agents who are located in the same node as they are (the community center).

The experimental scenarios would be:

1) World with the existing number of social workers.
2) A world in which an additional social worker is added to each community center.

3) A world in which 2 additional social workers are added to each community center.

7.3. COMMUNITY POLICING

Background:

Policing in Germany, and Berlin specifically, is based on an order-maintenance approach which is quite typical of European police more generally. Survey results indicate that Neukolln residents report a relatively high rate of negative interactions with the police. Overall, more than half of both Muslim and non-Muslim immigrants in Berlin have reported negative feelings or views of the police. According to the results of T2.1, negative interactions with the police is an important risk factor for radicalization.

At the same time, there is strong empirical evidence that community policing leads to more positive attitudes toward the police (Weisburd & Majmundar, 2019). We assume here that community policing in our model will accordingly have a positive impact on possible radicalization and recruitment through their encouragement of procedural justice (Tyler, Schulhofer and Huq, 2010). Tyler, 2012; Cherney & Murphy, 2013). In this scenario the police will carry out a special training course on procedural justice and community policing for a select number of officers. The course will occur before the experiment.

The experiment will include three scenarios:

1) The existing world of policing in the simulation
2) Each of the communities gets 1 additional community police officer,
3) Each of the communities gets 2 community police officers.

The effects of the community police officer are that s/he produces a positive effect on the legitimacy/trust opinion dynamic. Community police officers do not produce negative outcomes and 90% of interactions between them and citizens are positive. Community police officers only affect opinions on legitimacy/trust and integration and not on subjective deprivation.

The movement of community police officers is that at each tick of the model they will move to a new patch. They can only move to patches where at least 1 citizen agent is located at that time. They will remain at a given patch for one hour, so that they will traverse 1 patch per tick of the model.

The community police officer's effects on opinion dynamics will be limited to 4 citizen agents located at the same patch. This means that if only 1-3 agents are located on the patch, they will all be influenced by the officer. If 5+ agents are located at the same patch, only 4 of them will be influenced by the officer. The selection of which agents will and won’t be influenced in such a situation will be random.

8. References


Section 3 Reports on experiments (task T4.3)

Due to the decision of to develop separate experiments for organised crime and terrorist network Section 3 is divided into two parts:

- An Experiment on the Individual and Strategic Determinants of Criminal Collaboration
- Examining the Interactive Effects of Personalization Algorithms (the Filter Bubble) on Network Structure (the Echo Chamber) and the Impact on Radical Beliefs
An Experiment on the Individual and Strategic Determinants of Criminal Collaboration

Authors: CNR, LUISS

1. Summary

This report describes the experiment conducted to investigate and fill in some gaps for the organised crime network (OCN) agent-based model. Using an ethical substitute, the well-established, die-rolling task (Fischbacher & Föllmi-Heusi, 2013), we explore the individual and strategic determinants of collaborative “criminality”. Specifically, the experiment provides information on (i) how past experience and risk aversion are related to collaboration, (ii) whether individual determinants of immoral behaviour (dishonesty) is associated with determinants when immoral behaviour is collaborative, and (iii) whether experience and reputation have similar or differing effect on the collaborative criminality dynamic. We use the results derived from our experiment to calibrate the OCN agent-based and to test the social network focus of the model.

2. Introduction

This report presents the laboratory experiment designed and conducted under Task 4.3. The aim of this experiment is to supplement prior empirical research and fill gaps that are relevant for the OCN agent-based model (T5.1). In particular, the literature review identified that there is a gap in the quantitative predictors for the decision to engage in criminal collaboration (see D1.1). Additionally, there is little knowledge about who engages in criminal collaboration and with whom. To fill this gap, we conduct a laboratory experiment that investigates the individual and strategic determinants of collaboration in crime. We investigate who engages in collaborative dishonesty and whom these people choose as their collaborative partners. This experiment provides data for calibrating the OCN model.

Naturally, we are unable to use actual crime as our outcome in this study. Instead, we use an ethically acceptable substitute: the well-established “die-rolling” task (Fischbacher & Föllmi-Heusi, 2013). Subjects roll a six-sided die in private, and, report the number that they rolled. The higher the number on their die, the more they earn. However, since this is done in private, they can report
whatever number they wish, and, since the task is anonymous, this gives subjects the freedom to do so. The only factor holding people back from lying is their own internal motivations.

While not perfect, we believe that this die-rolling task captures an important part of the decision to engage in the kinds of criminality that many OCN members undertake: providing protection, racketeering, acting as intermediaries, and selling stolen or illegal goods. In contrast to types of crimes that are frequently done under altered mental states (e.g. homicide or terrorism), often, racketeering and organised crime is done from a more calculated perspective. Like in the die-roll task, the decision is largely a cognitive one. In addition, undertaking these crimes requires that OCN members break moral or social norms. Stealing, cheating, and intimidation are broadly considered to be normatively bad behaviours in society. Likewise dishonesty in the die-rolling task.

This experiment is also informative in two further ways. First, we use it to understand the effects of networks on dishonesty. Using a within-subjects treatment we compare how individuals behave when they have the opportunity to be dishonest in an individual setting compared to when they have the opportunity to be collaboratively dishonest. Second, we compare the role of experience and reputation in influencing collaborative dishonesty. This test can be used for future extensions of the OCN model.

The computerized experiment was run, using z-Tree software (Fischbacher, 2007), in LUISS CESARE Lab (Rome) with student participants recruited with ORSEE (Greiner, 2015). The participants belong to undergraduate or postgraduate programs in one of three faculties: Economics, Law, or Political Science.

Our experiment consists of four phases. In the first two phases, we use multiple methods to elicit subjects’ individual honest behaviour. In the third phase, we allow subjects to collaboratively undertake dishonest behaviour. Comparing behaviour in the third phase with the two preceding ones gives us a within-subjects manipulation1. In the final phase, subjects answer a questionnaire in which we elicit their demographics, subjective and objective trust measures, risk attitudes, cognitive reflection capabilities, and personality characteristics.

We also conducted two between-subjects treatments2: an Experience Treatment (ET) and a Experience and Reputation Treatment (ERT). In the former, subjects

---

1 Within-subjects design is an experimental protocol where the same individual tests all the conditions, i.e. in this case all participants participate in all three Phases of the experiment.

2 Between-subjects design is an experimental protocol where different individuals test different conditions, i.e. in this case some participants participate in all three Phases of the Experience Treatment while other participants participate in all three Phases of the Experience and Reputation Treatment.
only had the opportunity to build knowledge about the others in their group through direct experience while in the latter each also imperfectly revealed their past behaviour to everyone in their group. This way, subjects could tell how everyone in their group behaved in every period.

In summary, our laboratory experiment contributes to the OCN model in the following ways:

1) It tests the conceptual focus of the OCN model: the influence of networks on criminal collaborative behaviour.
2) Uses past experience and risk aversion to calibrate the propensity to dishonestly collaborate.
3) Tests the effects of reputation on the propensity to dishonestly collaborate, as well as on behaviour towards a collaborative partner.

In following sections, we describe the experimental setting and the procedure we use, our research questions, display our results, and outline our conclusions.

3. Structure and research questions

As mentioned in the previous section, one of the aims of this study is to test, in an experimental setting, when and how participants are willing to accept the cost associated with violating moral or legal norms in establishing immoral collaborative relations with the others. These relations could end with mutual benefit, but also in a loss to one or both partners, depending on their own as well as their partners’ decisions.

The decision to establish immoral collaborative behaviour could be exemplified in the real world by a situation where, in a given population, an individual can decide either to earn a living through legal occupation or through committing crime with others. We assume that in order for a criminal collaboration to start, individuals have to offer the opportunity to commit crime to others so that they can conduct small illegal ventures (in the real world: racketeering, money laundering, robbery, etc.) with a gain that is crucially related to the actual behaviour of their partners. This, however, is risky business with unpredictable outcomes, whereas earning living through legal occupation procures a comparatively safe and stable income.

Immoral behaviour and the individual propensity to break a social rule may be related not only to individual characteristics perceptions of social or legal norms, but also to the different pressure of living in a given environment where other’s behaviour can affect one’s individual decision-making. Moreover, since humans frequently cooperate in order to achieve profitable goals, this could also lead to joint rule violations. Obviously, reputational concerns are involved in these kind of decisions as well.

With this study, we investigate how such collaboration emerges and spreads depending on one’s network, and how this is affected by whether participants
can openly observe the surrounding, through reputation, or not. We are also interested in understanding if sharing responsibility matters in this setting. That is, are people more prone to avoid imposing profits on their partners, forcing them to become “partners in crime”? Or does collaboration have a liberating effect, “freeing” people to behave unethically?

We design an experimental setting that enables us to answer the following questions:

- How does the information received about the activity of others affect one’s own decision to enter an immoral activity?
- Do participants who are willing to enter immoral activity with a partner then end up betraying their partners?
- How do participants make their decision to team up in immoral behaviour?
- Which participants attempt to start immoral collaboration? Are there demographic characteristics that are associated with collaboration?

4. Experimental procedure

Since the more frequent real-world violations of rules and criminal behaviours cannot be implemented in the lab, we adopted the possibility to lie in order to maximize own payoff as the task mimicking the basic violation of moral rules. We adapted to our research question the established die-roll task (Fischbacher & Föllmi-Heusi, 2013), a task used for individual decisions, in a way that allows collaborative dishonesty (see Weisel & Shalvi, 2015).

All the participants in our experiment played the following sequence:

- the die-rolling task (Phase 1);
- the sender-receiver task (Phase 2);
- collaborative die-rolling task with endogenous partner choice (Phase 3); and
- a post-experimental questionnaire.

Our subjects are mainly university students. The payment for subjects is calculated based on their decisions, and the decisions of others, in the first three phases of the experiments. This is the sum of the payments of three randomly selected rounds, one for each phase. All subjects also receive a participation fee of €4. In the next section we present the experimental tasks as it was implemented.

4.1. START OF THE EXPERIMENT

At the start of the experiment the computer randomly forms groups of six participants. Every participant in each group is identified with a different symbol (Figure 9). Every participant is informed of her own symbol on her individual computer screen. This symbol identifies each participant for the entire experiment.

---

3 Since we always have more than one group of participants in the experimental laboratory, (three or four groups of six), it was impossible for subjects to recognize the other participants in her group even if they
4.2. **Phase 1: Individual Die-Roll Task**

In Phase 1 of the experiment, participants repeatedly play (10 times), the die rolling task (Fischbacher & Föllmi-Heusi, 2013). In this task, each subject rolls a six-sided die on her desk and then decides what number to report using her computer screen. The higher the reported roll, the more money she receives (Figure 11). Given that the true die roll is not monitored, subjects are free to report any number from 1 to 6, and may therefore report the truth or lie. As participants are paid based on what they report, lying can increase their earnings.

---

discussed the experiment after their participation because they would not know to which group they belonged. This feature of the experiment allows us to guarantee subjects’ anonymity.

*English translations for each of the screens is provided in the footnotes. “Groups and symbol. The computer has randomly generated groups of six participants. Group composition will remain constant throughout the whole experiment. Each group member is identified by a symbol, which has been randomly assigned. This is the symbol that was assigned to you, it will remain the same for the whole experiment.*
The aim of this task is to measure participants’ propensity to lie before the group task is played later on in the experiment. This allows us to separate individual propensity for lying as opposed to lying due to group effects. Since die-rolls are undertaken privately, we—the experimenters—never know if a specific individual is lying or not. However, we can—on the basis of simple probabilities—identify whether as a group subjects are honest. And, more importantly, since subjects repeat the die-rolling task 10 times, we can also create an individual-level honesty index that is fairly precise. This works on the basis of probabilities also; for instance, if a subject reports 10 rolls of six in a row, we can be fairly certain that at least for one of these rolls the subject is lying since the probability of obtaining ten rolls of six in a row honestly is miniscule (1.65 x 10⁻⁸).

Subjects are only told about their Phase 1 payment at the end of the experiment, after Phase 3, so as to not affect the future phases. They, however, are aware that they will be paid for one randomly drawn round in Phase 1 (from the 10 rounds) and that their payment will correspond to what they decided to report on their screen in that one randomly selected round (Figure 12).

---

5 “Phase 1: die rolling. Roll the die and report the outcome in the following box. Your payoff for this round will be equal to the reported number.”
4.3. **Phase 2: Sender-Receiver Task**

The second phase implements another measure of (dis)honest behaviour that is known as the sender-receiver task (Gneezy, Rockenbach, & Serra-Garcia, 2013). In the die-roll task, the person that dishonesty hurts is the experimenter in that reporting a higher roll spends money of the experiment. In contrast, in the sender-receiver task, dishonesty harms another subject in the experiment. Additionally, because of the way the task is implemented, we are able to infer at the individual-level whether subjects were honest or not.

While this extra precision implies that subjects’ decisions are not truly anonymous, it is the combination of the die-roll task (which is anonymous but less precise) and the sender-receiver task (which is not anonymous but is more precise) that we leverage to disentangle subjects’ propensities for dishonesty when alone and when in a group. The sender-receiver task is as follows.

The computer begins by randomly pairing the six participants from the same group and assigning each member of the pair either the role of the “sender” (mittente) or the “receiver” (ricevente) (Figure 13, Figure 14).

---

6 “Phase 1: payoff. Your payoff for this round is equal to 5.”
“Phase 2: pairs and roles. The computer has randomly matched you with another participant in the experiment. You have been randomly assigned the role of RECEIVER. The other member of the pair has been assigned the role of SENDER. Final payoff for this round will depend on both pair members’ choices.”
Then the computer presents on each participant screen two distinct tables of payoffs that represent a different distribution of payoffs between the two members of the couple, of which only one will be randomly selected and actually played (Figure 15).

---

8 “Phase 2: pairs and roles. The computer has randomly matched you with another participant in the experiment. You have been randomly assigned the role of SENDER. The other member of the pair has been assigned the role of RECEIVER. Final payoff for this round will depend on both pair members’ choices.”
Once the random draw occurs, the result is revealed only to the sender, who has the possibility to decide what to report to her partner about the random extraction (i.e. to lie to her or not) (Figure 16).

---

9 “Phase 2: payoff tables. Please observe carefully the two payoff tables proposed by the computer. The computer will randomly select one of them. Being a SENDER, you will be informed about which table is selected.”
After receiving the information form the sender, the receiver takes her decision (either U or D) (Figure 17). From this choice we can infer whether the receiver decided to trust the reporting of the proposer or not.

---

10 “Phase 2: table selection. The computer has randomly selected the following table. Choose which table you want to report to the RECEIVER as the result of the selection.”
The computer screen, again, will not show the final table depicting both choices of the sender and the receiver (and therefore the payoffs) to avoid the payment to influence their actions in Phase 3.

4.4. **Phase 3: Collaborative Die-rolling**

The collaborative die-roll task that we implement in Phase 3 of the experiment is composed of 30 rounds. In every round, participants, as originally split into groups of six, can partner with another participant from their group to engage in the task. Alternatively, participants can stay out in which case the participant receives a fixed earning of €2 and waits until the following round starts. If a subject decides to enter into collaboration, the computer matches participants on the basis of their preferences and the preferences of the other members in that group.

At the start of each round, from every subject we elicit their preferred matching rank (i.e. they have to order their potential partners on the basis of their preferences to be matched with). We allow the possibility to consider two or more equally suitable participants in the same group, and the possibility to not be matched with a certain subject.

All participants show their matching preferences by selecting at least one other group member for potential collaboration. They are then asked to indicate in a 11 “Phase 2: action. The other group member reported the following table as the one randomly selected by the computer. Choose which action you want to take on the basis of such reported table. ”
binary choice whether they wish to enter collaboration or not. In other words, whether to put their preference ranking “into action” or to leave the preference ranking dormant. The reason for this two-step process is a methodological one: it allows us to obtain information on people’s preferences even if they indicate that they do not want to actually collaborate. So, we learn whom they would have chosen had they decided to collaborate. Then, a random order is selected of the participants who opted in each round, and the participant who is drawn as first, gets to collaborate with the first person on her ranked list (given that this person also decided to opt in for collaboration, otherwise the second ranked participant is attempted etc.), followed by the second drawn participant who decided to go in, and so on.

Once pairing occurs, one random member of each pair is assigned the role of first mover, and the other of a second mover. The paired subjects interact in the modified version of the die-roll task (Weisel & Shalvi, 2015) by rolling the die and deciding what number to report to the other. The first-mover reports a die roll number, this reported number is shown to the second-mover, after which the second-mover also has to report his number.

The individual payment of each round for this phase depends both on individual’s and their partner’s reported rolls. In particular, the pair receives the reported roll of the die as payoff if they both report rolling the same number (e.g. if in a round they both report a 6 they receive 6 Experimental Currency Units (ECU)\(^{12}\) each; if they both report a 5 they receive 5 ECU each... , etc.). If the pair reporting differs by only 1, the one who reported the lower number receives the sum of the two reported die (e.g. if one reports 6 and the other 5, the one who reported 5 gets 11 ECU and the other nothing). If the reported die roll differs by more than 1, both players get 0.

This payoff setup models a risk to collaboration for the first-mover. By engaging in collaboration, the first-mover leaves open the possibility that the second-mover takes all of the “criminal” proceeds. By reporting a unit less than the first-mover, the second-mover can take the entire “pot”, i.e the second mover, in this case, will receive the sum of the two numbers.

In case collaboration does not occur, either because one decided to stay out or an odd amount of participants opt in, the non-paired subjects receive a flat rate of 2 ECU and wait until the following round starts. This flat rate is lower than the maximum that participants can earn if they engage collaborative die-rolling, but is higher than the expected payoff of the collaborative task if everyone would report the actual (honest) die roll.

---

\(^{12}\)The experimental currency units (ECU) are converted to Euros, at a predefined rate, at the end of the experiment.
At the end of each round the computer informs each participant about her payoff in that round. This information varies through the different treatments: in the Experience Treatment the computer informs participants about the number of the die roll reported by her partner, whereas in the Experience and Reputation Treatment the computer informs them about the number of the die roll reported by all the 6 members of the group who decided and succeeded in entering collaboration.

At the end of the experiment the computer randomly selects for payment one of the 30 rounds and the individual payoff for this phase corresponds to the payoff of that round.

4.5. QUESTIONNAIRE

The aim of the non-incentivized questionnaire administered at the end of the experiment is to elicit some demographics that could affect participants’ behaviour in the previous three phases. Participants should indicate first their gender, age, and level of studies. Subsequently they are submitted to several tasks.

In particular, individual risk attitude has been elicited through the traditional Holt and Laury test (2002). Trust and reciprocity attitudes have been detected presenting the traditional investment game (Berg, Dickhaut, & McCabe, 1995). Cognitive capabilities have been checked through the presentation of the questions (Frederick, 2005) the Cognitive reflection test (CRT). Personality traits have been measured by 10-item of the Big-Five inventory covering Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (Rammstedt & John, 2007).

4.6. TREATMENTS AND SESSIONS

We implemented two separate treatments of the experiment: Experience Treatment (ET) and the Experience and Reputation Treatment (ERT). We ran six sessions giving us, in total, 120 participants. Each session lasted around two hours. Different group of subjects participate to each treatment and nobody participate in more than one session. The average payment for each participant has been €18 euros including a participation fee of €4.

5. Results

Table 7 presents the frequency of the reported dice rolls in Phase 1, and is the first indication we find that participants are not honestly reporting their rolls. Of the 1200 dice rolls performed in Phase 1, if participants reported them honestly, we would expect to see that each number from 1 to 6 was reported roughly 200 times. Instead, we find that the numbers 1, 2, and 3 are reported much less frequently than this, and 5 and 6 are reported much more frequently. In fact, and in line with previous research (Fischbacher & Föllmi-Heusi, 2013), 6 is the
most frequently reported number representing virtually a third of all observations.

Table 7 – Frequency of reported die rolls in Phase 1

<table>
<thead>
<tr>
<th>Reported die roll</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>87</td>
<td>7.25%</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
<td>9.17%</td>
</tr>
<tr>
<td>3</td>
<td>145</td>
<td>12.08%</td>
</tr>
<tr>
<td>4</td>
<td>205</td>
<td>17.08%</td>
</tr>
<tr>
<td>5</td>
<td>259</td>
<td>21.58%</td>
</tr>
<tr>
<td>6</td>
<td>394</td>
<td>32.83%</td>
</tr>
</tbody>
</table>

Next, we look at the frequency of misreporting in Phase 2. While misreporting in Phase 1 is cheating and therefore immoral behaviour, participants may not perceive that they are hurting anybody by doing it. In fact, by cheating in Phase 1 participants are not hurting the other players in the experiment, but only the experimenters.

Misreporting in Phase 2, however, directly hurts another participant and may therefore be considered differently by them. Table 8 reports the frequency of misreporting by senders in Phase 2. It shows that only 50% of senders told the truth about which matrix box was randomly selected by the computer, therefore indicating that half of our participants lie even when this hurts other players, in this case their randomly drawn partner.

Table 8 – Frequency of lies told by senders in Phase 2

<table>
<thead>
<tr>
<th>Decisions of senders in Phase 2</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tell truth</td>
<td>30</td>
<td>50%</td>
</tr>
<tr>
<td>Tell lie</td>
<td>30</td>
<td>50%</td>
</tr>
</tbody>
</table>

Table 9 reports how many receivers played according to their best interest given the information they received from senders, i.e. a proxy for measuring how much they trusted the information that the senders sent them. This reveals that receivers expected senders to lie less than they actually did.

Table 9 – Receivers’ decisions in Phase 2

<table>
<thead>
<tr>
<th>Decisions of receivers in Phase 2</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ignored advice</td>
<td>19</td>
<td>31.67%</td>
</tr>
<tr>
<td>Followed advice</td>
<td>41</td>
<td>68.33%</td>
</tr>
</tbody>
</table>
5.1. OPTING INTO COLLABORATION

The payoff of the main phase of the experiment, Phase 3, was designed in such a way that opting into collaboration (i.e. forgoing the fixed payment and rolling the dice with a partner) is only worth it if one plans to cheat the experimenter by lying about the roll of the die. If one would stay out of the collaboration, they would receive a fixed fee of 2 euros, therefore the expected payoff of staying out is 2 euros. If one would go in, find a partner, and both players tell the truth, their expected payoff would be 1.55 euros.

Given the lower expected payoffs of opting into collaboration if one would tell the truth, we are particularly interested in how often this decision was made, and how it differs between the two treatments. In the ERT, from period 2 on, participants can see which of the other 5 members of the group opted into collaboration in the last round and what they reported as their die roll. In the ET, between rounds they can only see the die roll of their collaborating partner if they themselves opted into collaboration. Therefore, in the ET it is harder to deduce if and who exactly goes in and how they behave once they enter.

Figure 18 and Figure 19 show the frequency of the decision to opt into collaboration, and Figure 20 and Table 10 the frequency that one successfully enters (a match has between made between participants who opted into collaboration). While the decision to go in is prevalent, the rate at which it is chosen differs between the two treatments. Participants in ERT opt for this significantly more often than those in the ET. This finding is supported by the panel-logit regression shown in Table 11. Over the 30 periods that Phase 3 is played, this difference between the treatments increases rather than decrease (Figure 19). While the trend in ERT seems to be relatively constant, the frequency of opting into a potentially criminal collaboration in the ET shows a negative trend.
Figure 18 – Proportion of decisions to opt into collaboration by treatment

![Opted into Collaboration By Treatment](image)

Figure 19 – Opting into collaboration according to treatment and period

![Opted into collaboration By Treatment](image)
The regression in Table 11 shows an additional noteworthy finding. This regression examines the predictors to subjects decision to collaborate or not (record that this is done in a second separate step after they have indicated their preferred ranking of partners). The last six rows show that the decisions that a participant made in phases 1 and 2, do not lead to significantly different behaviour regarding entering into collaboration. All that seems to matter is whether the information they receive between rounds informs participants on
the reported dice roll of all collaborative members of the entire group or just their partner provided they had one (i.e. the treatment effect), and to a lesser degree their risk aversion.

*Table 11 – Deciding to opt into collaboration in Phase 3*

<table>
<thead>
<tr>
<th>Decision to opt into collaboration in Phase 3</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>-0.015***</td>
<td>-0.015***</td>
<td>-0.015***</td>
<td>-0.020***</td>
</tr>
<tr>
<td>Ref. cat. Experience and Reputation Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience Treatment</td>
<td>-1.358***</td>
<td>-1.365***</td>
<td>-1.392***</td>
<td>-1.386***</td>
</tr>
<tr>
<td>Ref. cat. 1 Collaborator ranked</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranked 2 collaborators</td>
<td>0.206</td>
<td>0.207</td>
<td>0.207</td>
<td></td>
</tr>
<tr>
<td>Ranked 3 collaborators</td>
<td>0.480*</td>
<td>0.481*</td>
<td>0.479*</td>
<td></td>
</tr>
<tr>
<td>Ranked 4 collaborators</td>
<td>1.005***</td>
<td>1.004***</td>
<td>0.998***</td>
<td></td>
</tr>
<tr>
<td>Ranked 5 collaborators</td>
<td>0.547***</td>
<td>0.550***</td>
<td>0.542***</td>
<td></td>
</tr>
<tr>
<td>Lie in Phase 2</td>
<td>-0.239</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role in Phase 2</td>
<td>0.184</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean reported Roll in Phase 1</td>
<td>0.115</td>
<td>0.105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ref. cat. Truth in Phase 2*sender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truth in Phase 2*receiver</td>
<td>0.458</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lie in Phase 2*receiver</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lie in Phase 2*receiver</td>
<td>-0.090</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.682**</td>
<td>2.758**</td>
<td>3.386***</td>
<td>4.010***</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

We next look at how frequently participants end up partnering with the same person of the previous round, based on the outcome of that round. We anticipate that having a successful previous collaboration is associated with a higher probability of future matching. Table 12 informs us that reporting the same number by both matched participants leads to higher frequency of repeated partnerships, which is much stronger when reputation is available within rounds.
Table 12 – Percentage repeated matching based on previous period outcome

<table>
<thead>
<tr>
<th>Repeated matching based on outcome</th>
<th>Experience and Reputation Treatment</th>
<th>Experience Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched at 6</td>
<td>71.73%</td>
<td>38.30%</td>
</tr>
<tr>
<td>Matched at 5</td>
<td>62.12%</td>
<td>45.16%</td>
</tr>
<tr>
<td>Matched at 4</td>
<td>53.06%</td>
<td>47.37%</td>
</tr>
<tr>
<td>Matched at 3</td>
<td>47.62%</td>
<td>17.37%</td>
</tr>
<tr>
<td>Matched at 2</td>
<td>61.56%</td>
<td>25.53%</td>
</tr>
<tr>
<td>Matched at 1</td>
<td>22.67%</td>
<td>18.02%</td>
</tr>
<tr>
<td>Undercut partner</td>
<td>25.30%</td>
<td>17.95%</td>
</tr>
<tr>
<td>Was undercut by partner</td>
<td>25.10%</td>
<td>17.72%</td>
</tr>
<tr>
<td>Else</td>
<td>21.88%</td>
<td>22.02%</td>
</tr>
</tbody>
</table>

5.2. **REPORTED ROLLS WHEN COLLABORATING**

Looking at the behaviour of participants once they have opted into collaboration,
Figure 21 shows the frequency of reported rolls separated by treatment and by whether the participant is a first or second mover. While the sequence of the individual who reports has a small effect on the reported dice roll, the difference between the treatments is striking. Participants in the ET report rolls very close to what one would expect to find if rolls of a die were reported truthfully, while participants’ choices in the ERT are very heavily skewed towards reporting fives and sixes. In fact, more than 65% of all reported rolls in this treatment are a 5 or 6, while only around 10% report rolling a 1 or 2.
Figure 22 – Reported rolls under collaboration by treatment and role

Figure 22 shows the mean reported roll in both treatments split by first and second movers over the 30 periods. Over time, just like in the decision to opt in, we see that the treatment differences increase rather than decrease. Participants in ET are consistently around, even slightly below, the expected “truthful” mean, whereas participants in ERT reach an average reported role of 5 in the second half of the phase, and remain there until end-game effects kick in. This difference is supported by the panel-tobit regression in Table 13, which finds that participants in ET report significantly lower rolls than those in ERT. Again, we find that behaviour in phases 1 and 2 is not indicative of different lying behaviour in Phase 3.
**Input (rules/specifications) for PROTON simulations and Wizard**

Figure 22 – Reported rolls under collaboration by period, treatment, and role

![Mean Reported Roll Graph](image)

Table 13 – Phase 3 reported die rolls

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Period</strong></td>
<td>0.026***</td>
<td>0.026***</td>
<td>0.026***</td>
<td>0.047***</td>
</tr>
<tr>
<td></td>
<td>(3.49)</td>
<td>(3.49)</td>
<td>(3.49)</td>
<td>(3.49)</td>
</tr>
<tr>
<td><strong>Ref. cat. Experience and Reputation Treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience Treatment</td>
<td>-2.037***</td>
<td>-2.036***</td>
<td>-2.098***</td>
<td>-2.421***</td>
</tr>
<tr>
<td></td>
<td>(-5.73)</td>
<td>(-5.73)</td>
<td>(-6.44)</td>
<td>(-6.38)</td>
</tr>
<tr>
<td><strong>Ref. cat. 1 collaborator ranked</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranked 2 collaborators</td>
<td>-3.004***</td>
<td>-3.003***</td>
<td>-3.006***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-6.15)</td>
<td>(-6.15)</td>
<td>(-6.16)</td>
<td></td>
</tr>
<tr>
<td>Ranked 3 collaborators</td>
<td>-2.407***</td>
<td>-2.406***</td>
<td>-2.412***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-6.04)</td>
<td>(-6.04)</td>
<td>(-6.07)</td>
<td></td>
</tr>
<tr>
<td>Ranked 4 collaborators</td>
<td>-3.059***</td>
<td>-3.058***</td>
<td>-3.072***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-7.92)</td>
<td>(-7.92)</td>
<td>(-7.97)</td>
<td></td>
</tr>
<tr>
<td>Ranked 5 collaborators</td>
<td>-3.945***</td>
<td>-3.945***</td>
<td>-3.961***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-12.12)</td>
<td>(-12.12)</td>
<td>(-12.20)</td>
<td></td>
</tr>
<tr>
<td><strong>Lie in Phase 2</strong></td>
<td>-0.168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.53)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Role in Phase 2</strong></td>
<td>0.027</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean reported roll in Phase 1</strong></td>
<td>0.133</td>
<td>0.134</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
<td>(0.65)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ref. cat. Truth in Phase 2*sender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truth in Phase 2*receiver</td>
<td>-0.037</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lie in Phase 2*sender</strong></td>
<td>-0.231</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 699824.
5.3. Behaviour of second movers

In this section, we analyse how second movers respond to the reported rolls of the first mover. By the time they are asked to report their own roll, second movers already know the reported roll of the first mover. Therefore, other than telling the truthful die roll, they have the opportunity to report the same roll as the first mover they were matched with and give themselves and the first mover a positive, and possibly high, payoff. Additionally, they can decide to undercut the first mover’s reported die roll by exactly one unit and receive an even higher payoff. Of course, both of these situations can happen even if the second mover reports the truthful roll, but the chances are one in six for the matching and even less for the undercutting, since a reported roll of one cannot be undercut. This means that, if one reports the dice roll truthfully, both of these situations (matched and undercut rolls) should sum up to less than a third of all reported rolls.

Figure 23 and Figure 24 indicate that of all reported rolls by second movers around 90% in ERT and 75% in ET either match or undercut the first mover by one. Interestingly in ERT, likely due to reputation, the majority of reported rolls matches the first mover, whereas in ET matched and undercut rolls are reported with similar frequency.
Input (rules/specifications) for PROTON simulations and Wizard

Figure 23 – Second mover die roll in the ET

Table 14 shows what affects the decision of second movers to undercut their partnered first movers, using a panel logit-regression. Looking at the second
row of the Table 14, we find that participants in ET more frequently undercut than those in the ERT, which is likely due to three factors. The first is that in ET it is much harder to be punished in future rounds for undercutting, the second is that in ET to find a participant to collaborate with (by both reporting high rolls) is harder. In fact, Figure 23 tells us that ET also has more second movers who neither undercut nor matched the first mover’s role, so likely reported the truth.

Table 14 – Decision by second mover to undercut first mover

<table>
<thead>
<tr>
<th>Decision by second mover to undercut first mover</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(-0.88)</td>
</tr>
<tr>
<td>Ref. cat. Experience and Reputation Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience Treatment</td>
<td>0.327**</td>
<td>0.324**</td>
<td>0.212</td>
<td>0.275*</td>
</tr>
<tr>
<td></td>
<td>(2.15)</td>
<td>(2.14)</td>
<td>(1.46)</td>
<td>(1.80)</td>
</tr>
<tr>
<td>Risk Aversion</td>
<td>-0.0975**</td>
<td>-0.094**</td>
<td>-0.105**</td>
<td>-0.115**</td>
</tr>
<tr>
<td></td>
<td>(-2.04)</td>
<td>(-1.98)</td>
<td>(-2.13)</td>
<td>(-2.22)</td>
</tr>
<tr>
<td>Ref. cat. 1 collaborator ranked</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranked 2 collaborators</td>
<td>-0.196</td>
<td>-0.202</td>
<td>-0.217</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.59)</td>
<td>(-0.61)</td>
<td>(-0.65)</td>
<td></td>
</tr>
<tr>
<td>Ranked 3 collaborators</td>
<td>0.151</td>
<td>0.151</td>
<td>0.151</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
<td>(0.65)</td>
<td>(0.64)</td>
<td></td>
</tr>
<tr>
<td>Ranked 4 collaborators</td>
<td>0.487**</td>
<td>0.487**</td>
<td>0.509**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.24)</td>
<td>(2.25)</td>
<td>(2.34)</td>
<td></td>
</tr>
<tr>
<td>Ranked 5 collaborators</td>
<td>0.561***</td>
<td>0.569***</td>
<td>0.532***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.16)</td>
<td>(3.21)</td>
<td>(2.99)</td>
<td></td>
</tr>
<tr>
<td>Lie in Phase 2</td>
<td>0.069</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role in Phase 2</td>
<td>-0.287**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean reported roll in Phase 1</td>
<td>0.148*</td>
<td>0.148*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.68)</td>
<td>(1.69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ref. cat. Truth in Phase 2*sender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truth in Phase 2*receiver</td>
<td></td>
<td></td>
<td>-0.174</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-0.94)</td>
<td></td>
</tr>
<tr>
<td>Lie in Phase 2*sender</td>
<td>0.179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.98)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lie in Phase 2*receiver</td>
<td>-0.220</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.93</td>
<td>-1.293**</td>
<td>-0.603**</td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td>(-1.62)</td>
<td>(-2.42)</td>
<td>(-2.06)</td>
<td>(-0.21)</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 14 also reveals that the decision by second movers to undercut the first mover is significantly and positively affected by the number of potential collaborators (group members) they ranked for collaboration. The more participants the second movers ranked, the more likely it is that they undercut the first dice roll. This indicates that some participants strategically ranked as many group members as possible in order to increase the chance that they enter
collaboration with anyone so that they can undercut them if they are the second mover. The decision to undercut the first mover is also negatively correlated with risk aversion, thus risk-loving participants are likely to betray their partners. Lastly, participants who reported higher rolls in Phase 1 are also more likely to undercut as second movers, though this result is only weakly significant.

Although a large percentage of second movers reported rolls that matched the rolls of first movers, this does not mean that they all matched on six—the most profitable roll. Table 15 shows the frequency of all prominent outcomes of participants who entered collaboration. As can be seen, a large number of participants matched on numbers other than six and very frequently on one. First movers may have chosen to report one if they have been undercut in previous rounds in order to punish or warn second movers.

Table 15 – Outcomes of matched partners in Phase 3

<table>
<thead>
<tr>
<th>Experience and Reputation Treatment</th>
<th>Frequency</th>
<th>Percent</th>
<th>Experience Treatment</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched at 6</td>
<td>774</td>
<td>42.43%</td>
<td></td>
<td>54</td>
<td>6.43%</td>
</tr>
<tr>
<td>Matched at 5</td>
<td>132</td>
<td>7.24%</td>
<td></td>
<td>34</td>
<td>4.05%</td>
</tr>
<tr>
<td>Matched at 4</td>
<td>50</td>
<td>2.74%</td>
<td></td>
<td>40</td>
<td>4.76%</td>
</tr>
<tr>
<td>Matched at 3</td>
<td>44</td>
<td>2.41%</td>
<td></td>
<td>24</td>
<td>2.86%</td>
</tr>
<tr>
<td>Matched at 2</td>
<td>26</td>
<td>1.43%</td>
<td></td>
<td>50</td>
<td>5.95%</td>
</tr>
<tr>
<td>Matched at 1</td>
<td>166</td>
<td>9.10%</td>
<td></td>
<td>136</td>
<td>16.19%</td>
</tr>
<tr>
<td>Undercut partner</td>
<td>280</td>
<td>15.35%</td>
<td></td>
<td>185</td>
<td>22.02%</td>
</tr>
<tr>
<td>Was undercut by partner</td>
<td>280</td>
<td>15.35%</td>
<td></td>
<td>185</td>
<td>22.02%</td>
</tr>
<tr>
<td>Else</td>
<td>72</td>
<td>3.95%</td>
<td></td>
<td>132</td>
<td>15.71%</td>
</tr>
</tbody>
</table>
Figure 25 and Figure 26 display the most prominent outcomes, over time, and depict the role reputation plays in collaborative dishonesty environments. The ability to see past reputation of a potential collaborative partner leads to much more stable and profitable collaboration. It also shows that the frequency of both players reporting six increases over time, before the end-game effect\textsuperscript{13} kicks in. Without it undercutting, reporting unequal dice, as well mutual reports of one, become much more prominent.

\textsuperscript{13} End game effect is a common phenomenon in games of cooperation, where participants stop cooperating in the last period because there is no future reputation to uphold, i.e. acting uncooperatively in the last period (or two) cannot hurt them (much) in the future.
6. Conclusions

A main result of our study is that individual and team dishonesty have very different roots. Individual dishonesty in Phase 1 and Phase 2 has little association
with collaborative dishonesty—we only found a barely significant association between Phase 1 decision and reported die-roll as second mover in Phase 3. Instead, what matters for reporting in Phase 3 is the length of exposure to the actions of others and the kind of information and experience that one receives. We also show that experience with partners, in both the ET and the ERT treatments, matters for whether future collaborations occur. A good experience, matching at six or a five for instance, is associated with a higher probability of repeated matching than a poor experience, for instance being undercut (Table 12). We use this result to inform the effect that previous experience has on the probability of future collaborations in the OCN agent-based model.

The comparison between the two different treatments implemented shows that engaging in a criminal activity is strongly related to the behaviour of subjects’ potential partners and on how much they are willing to collaborate with them. Relative to experience only, including reputation on other group member’s reporting behaviour leads to more dishonesty. In any case, the choice of a partner is crucial in this setting and reduces the possibility for betrayal to occur (i.e. in the ERT treatment the number of undercutting from the second mover is sensibly lower than in the ET treatment).

These results support the network focus of the OCN agent-based model. In our experiment “criminal” collaboration is, at least in part, driven by the social networks that individuals form and thus these need to be taken into account in models of OCN.

We also find that without reputation transmission there are fewer collaborations and that there is more honest reporting. But, that if subjects enter collaboration there is more undercutting—or betrayal of trust—leading to, on average, worse outcomes both for both partners. This suggests that while direct experience matters, reputation is also an important part of OCN formation in the real world. A potential future extension of the OCN agent-based model is to include reputation—a far from simple task—and explore how this shapes the model OCN formation.

Regarding individual characteristics, we show that risk aversion is negatively associated with collaboration (Table 11). People who are more risk averse are less likely to engage in collaborative die-rolling. We use the association between risk aversion and collaboration to inform the calibration of the OCN model, particularly in calibrating agent’s propensities. In future work, we plan to conduct further analyses that investigate the socio-economic and demographic predictors or collaboration, and, intend to include these factors into the OCN model.
7. References

Examining the Interactive Effects of Personalization Algorithms (the Filter Bubble) on Network Structure (the Echo Chamber) and the Impact on Radical Beliefs

Authors: Michael Wolfowicz, David Weisburd and Badi Hasisi
Partner: HUJI

1. Introduction

There is an ongoing debate as to whether the internet is in and of itself a radicalizing agent, an amplifier of radicalization, whether it is both, or whether it is neither (Neumann, 2013; O'Hara & Stevens, 2015). That the literature is lacking in quantitative inquiry has not helped to reconcile this debate (Gill et al., 2017). While the rise of internet usage may not correlate with an increase in terrorism (Benson, 2014), it is still possible that certain elements and aspects of the internet may be contributing to radicalization. In recent years, scholars have begun to focus on the role of so-called filter bubbles, echo chambers, and the potential for algorithmic deviancy amplification (Hawdon, 2012; Wood, 2016).

The research on online radicalization emphasizes that certain types of networks, characterized by high levels of insularity and density, are fundamental in
reinforcing radical beliefs. Dense and insular networks made up of highly similar individuals have come to be referred to as 'echo chambers'. Like offline, individuals traversing the online environment have a tendency to actively seek to be part of a social network of likeminded people (McPherson et al., 2001). However, certain environmental factors may also be complicit in determining associations and networks. With regards to the internet specifically, researchers have recently begun to examine whether personalization algorithms may be complicit in determining what content users will and will not see, and in shaping online networks and their structures. The role of personalization algorithms as an environmental factor, and network structures as a learning environment, have potentially important implications for the learning of deviant beliefs and behaviors, and radicalization specifically (Schmitt, Rieger, Rutkowski and Ernst, 2018).

The effects of personalization algorithms on contributing to the development of echo chamber type networks is known as the "filter bubble" effect (Pariser, 2011; Datta et al., 2015). If a user displays some interest in a particular subculture, even unintentionally, the filter bubble automatically provides them with similar content and associations of the same genre (Gottron & Schwagereit, 2016; Hawdon, 2012; Wood, 2016). One danger is that the filter bubble is somewhat of a black-box. It has been theorized that it may inadvertently bring radicals and other deviants into contact with each as well as reinforcing content, whilst shielding them from opposing views. The interplay between the filter bubble and echo chamber may carry the added risk of "algorithmic deviancy amplification" (Wood, 2016).

In this study we set out to examine the role of personalization algorithms in determining network structure characteristics and how network structure characteristics may affect the development of radical beliefs. We conducted a randomized control trial on a sample of new Twitter users from East Jerusalem. The participants were randomly assigned to a treatment group of algorithm suppression in which all personalization options were de-selected. We
hypothesize that treatment group's networks will be more externally focused and as a result, that the participants will be less likely to express radical beliefs.

2. Theoretical framework

2.1. Radicalization and Online Radicalization

The EU's definition of radicalization describes radicalization as "the phenomenon of people embracing opinions, views and ideas which could (sic.) lead to acts of terrorism" (EU, 2005). While this definition, and indeed most definitions of radicalization, does not specify what "opinions, views and ideas" it refers to, most scholars now agree that any measure of radical beliefs must include an element of support or justification for violence. When examining beliefs that relate to behaviors, it is important that the measure chosen for the belief has a high level of specificity with regards to the behavior of interest. As such, in examining the types of beliefs that "could lead to acts of terrorism", support or justification for terrorism represents the belief with the highest degree of specificity. Indeed, many studies now use measures of support or justification for radical violence, terrorism (generally or specific acts, e.g. 9/11, or 7/7 bombings), or suicide bombings as proxies for radicalization (e.g. Tausch, Spears and Christ, 2009; Victoroff, Adelman and Matthews, 2012; McCauley, 2012 Zhirkov, Verkuyten and Weesie, 2014; Berger, 2016).

This conceptualizing and approach to measuring radicalization can also be applied to online radicalization. According to Berminham (2009), online radicalization is "a process whereby individuals through their online interactions and exposures to various types of internet content, come to view violence as a legitimate method of solving social and political conflicts". While online radicalization is widely discussed, prior quantitative research on online radicalization has been somewhat sparse (Gill et al., 2017). Among those studies which do provide quantitative analysis, the focus is generally on
mediated exposure to violent content. These studies often combine measures of exposure to mediated violence from a number of sources, thereby making it difficult to disentangle the effects of the internet specifically (e.g. Nivette, Eisner and Ribeaud, 2017; Baier et al., 2010; Gvirsman et al., 2016). Among the few experimental studies that have been conducted, the simple exposure to radical content on support or justification of radical violence has been found to have little to no effect (Rieger et al., 2017; Kalmoe, 2016; Shorland et al., 2017).

While violent media more generally has long been found to affect violent cognitions and behaviors (Bandura, 1978), the internet differs from traditional media in that it allows for both passive and active forms of consumption (Pauwels & Schils, 2016). A series of recent studies examining Flemish youth differentiated between passive and active engagement with extremist content online and its effects on radical beliefs and well as self-reported radical behaviors. The studies found that the effects of active engagement are somewhat larger than passive exposure (Pauwels & Schils, 2016; Pauwels & Boudry, 2017; Pauwels & Svensson, 2017; Pauwels & Heylen, 2017). Other studies have found that the active posting of politically related content has a positive association with radical beliefs, including the justification of terrorism and support for foreign fighters (e.g. Wojcieszak, 2010; Bhui et al., 2016; Pedersen et al., 2018).

Prior studies have found that for the majority of terrorism offenders, the internet has played at least some role in their radicalization process. The primary areas in which the internet is used by these radicals is for: communications, reinforcing prior beliefs, seeking legitimization for action, disseminating and consuming propaganda, engaging with support groups, and planning (Gill et al., 2017). As such, while there has been some quantitative research on the dissemination and consumption of radical content, little work has been done on the role of networks, especially with regards to how they may reinforce radical beliefs.
2.2. **The Echo Chamber: The Importance of Network Level Characteristics**

Prior studies on radical online networks have found that these networks can be characterized by high density and insularity. While ties within the networks are often quite weak or loose, there is at the same time a high level of tie reciprocity and interconnectedness. Depending on their values across such factors, these networks may come to represent what Sunstein (2007) referred to as an "echo chamber". An echo chamber is a highly insular and inward focused network structure where messages in support of a particular belief or behavior receive constant reinforcement and opposing narratives hardly exist. Many scholars have suggested that echo chamber networks may be an appropriate framework for understanding the role of networks in radicalization (Stevens & Neumann, 2009; Sunstein, 2017; Sunstein, 2007; Von Behr et al., 2013; Warner, 2010; Wojcieszak, 2010; Gilbert et al., 2009; Del Vicario et al., 2016a, 2016b).

A highly insular and homogenous network is considered to be a prime "criminogenic environment" (Sutherland, 1947; Neumann, 2013; Von Behr et al., 2013). Indeed, some scholars have theorized that network structures and characteristics (such as density and insularity) may be more important to the development of deviant beliefs and behaviors that the identity of the network's members (Burt, 1992; Haynie, 2001; Haynie, 2002; McGloin & O'Neill Shermer, 2009; McGloin & Piquero, 2010). Similarly, with respect to radicalization, wider network structures may play a more important role than individual associations (Kennedy & Weimann, 2011; Malthaner & Waldmann 2014).
According to social learning theory perspectives, the learning of deviant behaviors occurs in the same way as the learning of normative behaviors, in intimate social networks. Given the individuality of networks, the learning of any behavior must be understood "in the context of all other concurrently available schedules and sources of reinforcement" (Akers, 1998:70). This means that the structure of the networks need to be examined as well. But social structures and the networks that make them up are dictated by a range of external and environmental factors (Matsueda, 1988), including online (Holt, Buruss & Bossler, 2010). In this context, personalization algorithms represent one of a many environmental factors that shape and determine online network organization and structures (Weisburd, 2010; Hawdon, 2012; Bessi et-al, 2016). It has previously been suggested that personalization algorithms contribute to the development of echo chamber type networks where the social learning of radical beliefs occurs and may even be amplified (Hawdon, 2012; Wood, 2016).

2.3. THE FILTER BUBBLE

The "filter bubble" is a term that refers to and describes the collective effects of personalization algorithms. According to the filter bubble hypothesis, these algorithms influence and dictate the content and associations that will be made available to an individual based on their online profiles and histories (Pariser, 2011). As one example, Facebook's algorithms “control the ‘visibility’ of friends, news, items, or ideas” (van Dijck, 2013:49). As a result, they determine not only differential associations but also the shaping of networks (Skeggs & Yuill, 2016).
While researchers and policy makers have discussed the potential dangers of the filter bubble extensively, surprisingly little is known about how the effects on content exposure and network development, as well as cognitive and behavioral outcomes. A study commissioned by Facebook found that the platform's algorithms resulted in a reduction in exposure to ideologically opposing content of 8% for Liberal users, and 5% for Conservative users. The study claimed that these reductions were effectively insignificant (Bakshy et-al., 2015). However, an independent study found that the reduction in exposure to cross-cutting ideological content was actually 25% for politically conservative users and a whole 50% for Liberal users (Nikolov et-al, 2015).

Other studies have examined the role of personalization algorithms with regards to their effects on exposure to radical content specifically. For example, analyzing Youtube's keyword matching system, Musa and Bendett (2010) found that even neutral searches on "Islam" resulted in recommendations for some radical Islamist content. The authors explain that by clicking on this content, the recommendations are further refined and within a few clicks a user can be immersed in such content (Musa & Bendett, 2010). Similar findings have been made for extreme right-wing content, which may be reached through non-extremist Youtube searches and Twitter links (O'Callaghan et-al, 2013, 2014). When Regner (2014) simulated the YouTube activities of left-wing and right-wing radical users, there was almost no opposing content, and violent content soon became part of the top recommendations and playlists.
In another study, Dandekar et-al (2012) found that two of the three personalization algorithms they tested had polarizing effects for biased users, while the other had polarizing effects even for unbiased users. While some experimental examinations of the effects of personalization on polarization have been carried out in different fields, the findings are mixed (e.g. Conover et-al, 2011, 2012; Nikolov et-al, 2015). With regards to more radical ideologies, a comparison of hardcore 'science' and 'conspiracy theory' networks found that nearly all members' networks were echo chambers. The researchers also noted that by a user's fiftieth action, the effects of personalization on polarization could already be observed (Bessi et-al, 2016).

As the studies above demonstrate, the current evidence is highly contradictory (Zuiderveen Borgesius et-al, 2016). Moreover, the effects of the filter bubble on network development and structure have not been studied in the way that content exposure to radical content has. There are therefore large gaps in this relatively small literature. While personalization algorithms have been theorized to increase the risk for exposure to radical content and thereby radicalization, there has to date been no study that examines this issue specifically, or which includes the mediating effects of network structure (Hawdon, Bernatzky, Costello, 2018; Schmitt, Rieger, Rutkowski and Ernst, 2018).

3. The current study

In this study we conduct a randomized controlled experiment to test the effects of personalization algorithms on the creation of echo chamber type networks,
and the effects of network structure characteristics on predicting radical beliefs. This means that we do not expect that the treatment will have any direct effect on radical beliefs. Rather, the interactions between the treatment and network structure characteristic variables are what are of prime concern.

3.1. TREATMENT AND SETTING
Given the nature of the study, it was necessary to create a treatment that rather than providing participants with radical content, would theoretically reduce the insularity, density, and inwardness of their networks (as well as affect other network structure characteristics), and thereby their exposure to one-sided content. The treatment in this study was therefore 'algorithm suppression'.

In order apply the treatment, participants in the treatment group were first asked to create new email addresses for signing up to Twitter. Upon signing up, participants from the treatment group skipped all automatic recommendations. Following this, and immediately upon opening their accounts, they turned off all personalization algorithms in the privacy and security settings and options. While participants were not given any specific tasks, they were encouraged to try utilizing Twitter as their primary social media platform during the experimental period.

Twitter was also chosen as the platform for the experiment since it is easier to find new users who do not have Twitter, as well as the ability to carry out SNA on Twitter data. Additionally, while Facebook remains the primary social networking platform, and while it is certainly popular for extremism, Twitter has arisen as a central platform for radical milieus. Moreover, echo chambers are
thought to be especially prevalent on Twitter compared to other social networking platforms.

4. Sample

Recruitment for this study was carried out based on responses to fliers and digital advertisements in East Jerusalem, including the Hebrew University of Jerusalem campus on Mount Scopus. In order to be accepted as a participant, applicants had to currently use Facebook but indicate that they had never have used Twitter. Participants also had to indicate that they actively used Facebook and other social media to access news and politically driven content. The sample included young adult males and females from East Jerusalem, a population that is considered to be at high risk for radicalization.

The original sample, which commenced the study by opening their Twitter account on a single day in January 2019, consisted of 115 participants, with 57 participants in the control group and 58 in the control group. Over the course of the next three months, 12 participants elected to withdraw from the study. Of the remaining 103 participants, 7 failed to complete the survey. As such, the final sample consisted of 96 participants, with a similar number of dropouts from each group the final size of the control group was 49 and for the treatment group 47. Overall, the sample was 84.4% female and the mean age was 18.66 (SD=2.025). The average level of education was 2.915 (SD=.964), indicating that the average participant was currently in university preparatory studies.
There were no statistically significant differences between the treatment and control groups on the demographic variables (See Table 1).

**Table 16: Descriptive statistics of sample demographics**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>19.021 (2.480)</td>
<td>18.306 (1.432)</td>
</tr>
<tr>
<td>% Female</td>
<td>85.1%</td>
<td>83.7%</td>
</tr>
<tr>
<td>Education</td>
<td>2.867 (1.036)</td>
<td>2.959 (.912)</td>
</tr>
</tbody>
</table>

5. **Analytic strategy**

The first stage of the analysis involved conducting Social Network Analysis (SNA) for each individual participant's profile. For this process we used the NodeXL software package. Each user's network was constructed by including followers and followed users, as well as interactions between users (e.g. Tweet and re-tweets). Network data was collected over the 4-month experimental period as well as at the end of the period. For each user a number of network structure level parameters were calculated.

At the end of the 4-month experimental period participants completed an anonymous online survey. The data were linked to the data from the SNA automatically through the use of randomly generated keys which had been distributed to the participants prior to the start of the survey. Participants were asked a number of questions pertaining to their online experience during the period, their beliefs about their networks, as well as their radical beliefs. The majority of the surveys were completed in full, with only a small number of participants failing to respond to a small number of items. These missing observations were subsequently imputed using nearest neighbor imputation.
We subsequently combined the SNA and survey data and first examined them using t tests. However, the small sample size meant that there was low statistical power and it would be unlikely to be able to identify small effect sizes using this method. Additionally, given that the study was interested primarily in examining interaction effects between the treatment and network structure characteristics on radical beliefs, such tests could not effectively answer our primary research questions. As such, we ran a series of ordinal logistic regression models in which we tested the independent effects of each of the variables, as well as the interaction effects between the treatment, network-level variables, and items from the survey. Given that we were not interested in the direct effects of the treatment, and in order to reduce multicollinearity, treatment was not included as a stand alone covariate in the models.

6. Variables and hypotheses

6.1. Dependent variable

Radicalization- Our proxy for radical beliefs is "justification of suicide bombings", which was measured on a 5-point Likert scale. Participants were asked whether they felt that suicide bombings could ever be justified, with 1 indicating that they are never justifiable and 5 indicating that they are often justifiable. Similar measures have been used by the PEW global attitudes survey and the European Values Survey which have been used in a number of studies (e.g. Victoroff, 2010; McCauley, 2012; Berger, 2016 etc.), and other studies...
have used justification of specific suicide bombing events have (e.g. Tausch, 2009; Tausch, 2011).

6.2. INDEPENDENT VARIABLES

*Echo chamber*- Our primary measure for echo chambers is Krackhardt and Stern’s “E-I” ratio (or E-I index) which measures the ratio between external and internal ties. This measure has already been used in a number of studies examining echo chambers (Bright, 2016; Hargittai et-al, 2008; Everton & Cunningham, 2015; Everton, Cunningham & Murphy, 2016). The measurement is calculated as:

\[ E - I \text{ Ratio} = \frac{Ge - Gi}{Ge + Gi} \]

In this equation Ge is the total number of external ties and Gi the total number of internal ties. The E-I ratio has a scale of -1 to +1, where negative scores represent an individual or network that is more internally oriented (Bright, 2016). Individuals with scores closer to 1 have more externally focussed networks and would be less likely to adopt deviant attitudes and behaviors.

**Hypothesis 1**: The treatment interacts to create larger E-I scores and reduces the likelihood of radical beliefs.

*Closeness centrality (CC)*-While there are a number of different measures of network centrality available in social network analysis, closeness centrality measures a node’s position as a potential influencer over the rest of the network. The measure is based on the node's closeness to all other nodes based on calculating the shortest paths between all nodes. Higher CC scores indicate that an individual user may be better positioned to be an influencer over the network.
In networks, opinion leaders may be under additional pressure to act in accordance with the way they are perceived (Oeldorf-Hirsch & Sundar, 2015).

**Hypothesis 2:** The treatment interacts to reduce network centrality and thereby the likelihood of radical beliefs.

*Network density-* Network density is considered to be a good predictor of the learning of deviant behaviors (Haynie, 2001, 2002; Haynie & Kreager, 2013; McGloin & O'Neill Shermer, 2009; Hargittai et-al, 2008). Network density is associated with elements of social control and denser networks encourage conformity (Krohn, 1986). Generally speaking, this means that denser network will be less prone to deviance. However, when networks are deviant in nature, and the deviant beliefs are the norm, increased density would increase social control over network members to conform to these beliefs.

**Hypothesis 3:** The treatment interacts with network density to reduce the likelihood of radical beliefs.

*Reciprocated vertices-* Reciprocity is a quantitative measure that measures the number of mutual, or returned ties. For example, if A is friends with B and B is also friends with A. On Twitter not all users followed will follow the follower back in return. Previous studies have found that online right-wing extremist networks are characterized by high reciprocity (Burris, Smith and Strahm, 2000). According to the Strength of Weak Ties Theory (Granovetter, 1983), higher reciprocity indicates a closer and more intense relationship. Weak ties however, provide individuals with what they perceive to be increased opportunities for community and acceptance. Previously, Kennedy and Weimann (2011)
suggested that larger networks made up of a larger number of weak ties actually increases the likelihood of radicalization and that most radical networks are characterized by weak ties.

**Hypothesis 4:** The treatment interacts with reciprocity and thereby affects the likelihood of radical beliefs.

### 7. Results

We ran a series of ordinal logistic regression models in which we tested the interaction terms between the treatment and network structure characteristics individually whilst controlling for all other network structure characteristics. In model 1 we show the simple relationship between treatment and radicalization. As expected, this relationship is small and not statistically significant. In models 2-6 we do not include the main effect of treatment because of concerns re multicollinearity. In model 2, the interaction between the treatment and the E-I ratio was marginally significant. In model 3, the interaction between the treatment and density was significant. In model 4, the interaction between the treatment and centrality was not statistically significant. In model 5, the interaction between the treatment and reciprocity was not statistically significant. In model 6, which included all four interaction terms, the interaction between the treatment and the E-I ratio was significant at the .01 level, whilst density was also marginally significant. This generally confirms our main hypotheses. Where density and the EI ratios are high, subjects that had personalization algorithms turned off were less likely to hold radical beliefs.
### Table 17: Ordinal logistic regression models predicting radical beliefs

<table>
<thead>
<tr>
<th>Factor</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>-.254 (.365)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-I ratio</td>
<td>-2.771 (1.262)*</td>
<td>-2.701 (1.257)*</td>
<td>-2.496 (1.184)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>-2.676 (1.921)</td>
<td>-2.701 (1.262)*</td>
<td>-2.496 (1.184)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centrality</td>
<td>1.816 (1.868)</td>
<td>.068 (.447)</td>
<td>2.001 (1.651)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reciprocity</td>
<td>-.851 (2.392)</td>
<td>-4.791 (1.991)*</td>
<td>-3.819 (1.929)*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment* E-I ratio</td>
<td>-.900 (.509)†</td>
<td></td>
<td></td>
<td>-1.807 (1.710)**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment* density</td>
<td>-.858 (1.135)*</td>
<td></td>
<td></td>
<td>-2.477 (1.521)†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment* centrality</td>
<td>.084 (.096)</td>
<td></td>
<td>2.089 (1.506)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment* reciprocity</td>
<td></td>
<td>-1.234 (1.398)</td>
<td>-1.214 (1.729)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$(Nagelkerke)</td>
<td>.306</td>
<td>.419</td>
<td>.415</td>
<td>.424</td>
<td>.428</td>
<td>.418</td>
</tr>
</tbody>
</table>

† <.10, * <.05, ** <.01

## 8. Discussion

Studies on online radicalization have generally done little to examine the differential effects of the different elements, features and activities that combine to make up the online experience. Rather than being a single issue, there are a number of elements that go in to online radicalization, and these factors may also serve to act as predictors of radicalization. In this study we set out to examine whether the echo chamber increases radical believes. Prior studies on online radicalization have overwhelmingly focussed on the aspect of passive exposure to violent and radical content. However, there is little evidence that mere exposure plays any significant role in radicalization on its own. Indeed, the findings of our additional models were that passive exposure to radical content had no statistically significant effect. In line with the echo chamber
hypothesis and social learning perspectives, network structure characteristics represent the fundamental environment in which the learning of any attitude and behavior takes place. With regards to deviant attitudes and behaviors, such as radicalization, networks that are denser, more insular, and more inwardly focused may represent echo chambers in which network members are at an increased risk for polarization, bias confirmation, and radicalization. This is directly consistent with our findings. While shutting off the personalization algorithms per se did not have significant impacts in our study, when this treatment was conditioned on network characteristics (density and EI) the effects were statistically significant.

The significant effect for the interaction between the treatment and the E-I ratio indicates that for users who have de-activated their personalization algorithms, and have higher E-I ratios, there is a significant decrease in the likelihood of radicalization. While we found no evidence for the effects of network centrality on radical beliefs, we found some evidence regarding the other two network structure characteristics of density and reciprocity.

Traditionally, scholars have characterized echo chambers as being highly dense networks. However, this perspective tends to stand in contradiction to perspectives that hold that denser networks actually have higher levels of social control and thereby likely to reduce the likelihood of various forms of deviance (Browning, Feinberg & Deitz, 2004). Indeed, this perspective has generally been supported by studies from criminology. Following from this, our study found that there was a significant interaction between the treatment and network density.
in reducing the likelihood of radical beliefs. Another reason that networks with higher density decrease the likelihood of deviance is because denser networks are the result of the presence of a larger proportion of strong ties. As discussed, stronger ties also increase levels of social control and reduce the likelihood of deviance.

As discussed above, according to the Strength of Weak Ties Theory (Granovetter, 1983), stronger ties are believed to increase social attachment and social control. Conversely, weak ties can influence network members by providing them with new information and the perceived opportunities for community and acceptance. It is on this basis that Kennedy and Weimann (2011) suggested that larger networks made up of a larger number of weak ties increases the likelihood of radicalization. Previous studies support this idea and have found that terrorist networks are often characterized by a large number of weak ties (Varanese, 2016). However, our findings indicate that there was no significant interaction with the treatment in this regard.

Relatedly, the reinforcement provided by networks represents a key characteristic of echo chambers that has its basis in social learning theory. According to social learning theory, both differential and vicarious reinforcement increase the likelihood that a deviant attitude or behavior will be adopted; and this holds true for radicalization as well (Akers & Silverman, 2004). Interestingly, Shapiro & Maras (2018) emphasize the role of differential reinforcement in contributing to the development of radical beliefs in females specifically. In connection with the role of weak ties and reciprocity, individuals
are likely to mimic the attitudes and behaviors which they view as generating acceptance and respect, and avoid those which they perceive the network as finding abhorrent. Indeed, one characteristic of echo chambers is the abhoring of contradictory beliefs and positions (Sunstein, 2007; Geeraerts, 2012; Hawdon, 2012).

While social learning theory often emphasizes the role of differential associations and reinforcement, extensions of the theory, such as Social Structure Social Learning theory insist that more attention needs to be paid to the social structures and networks in which learning takes place. Such a focus is in line with one of the underlying premise of social learning theory, that deviant beliefs and behaviors are learned in the same way as normative ones. While social structures and networks remain an under-researched area, a number of scholars have suggested that these elements of social learning may even be more important than the identity of the members in networks. In this regard, the echo chamber framework emphasizes the network structure and its characteristics over the actual learning processes. Rather, it acknowledges that normal learning processes occur in any network, but that networks characterized by high levels of insularity and inwardness increase the risk of the learning of deviant beliefs and behaviors.

From a policy perspective, our results indicate that the suppression of personalization algorithms can have an impact on the likelihood of the development of radical beliefs. That impact importantly, is conditioned by network structures. Policy makers have been discussing the regulation of
personalization algorithms now for some time (European Commission, 2012). It is important to remember however that while personalization algorithms and networks can increase ideological fragmentation, they also enable for the exposure to cross-cutting content that could serve as a counter-balance (Flaxman, Goel and Rao, 2016). Indeed, studies have found that in non-western countries, those who have internet access are less likely to support terrorism than those who don’t (Fair & Salva, 2019). As such, whilst we found evidence to support aspects of the filter bubble and echo chamber hypotheses, we would not suggest that that personalization algorithms are inherently negative. In fact, they could very well serve as the ideal vehicle for facilitating effective counter-messaging to combat radicalization online (Schmitt, Rieger, Rutkowski and Ernst, 2018).

9. Relevance to Agent Based Model

The agent based model includes internet usage/communications as a variable operating within the model. However, little is known about how to weight this form of interaction relative to offline interactions. The findings of the current study indicate that network density and network insularity/inwardness, as measured by the E-I ratio, have a significant impact on the likelihood of radical beliefs. Importantly, by removing personalization algorithms, individuals are limited to ties that are made purely by self-selection. When this self-selection results in denser and more externally focussed networks, there is a reduced likelihood of radicalization. This to some extent reflects the rules that govern...
the opinion-dynamic function that is being used to model offline interaction effects and justifies using it to model online communications as well.

10. Conclusions

The current study is one of a handful of randomized control experiments that have been conducted in radicalization research to date, and with respect to online radicalization in particular. The study's findings find partial support for both the filter bubble and echo chamber hypothesis. Most importantly, our study shows that the impact of the personalization algorithms may be conditioned by network characteristics, namely the echo chamber. This provides not only important new knowledge, but also suggests that interventions should be focused on those with looser networks made up of weak ties, and more inwardly focused network characteristics. While we did not find that the treatment had a significant interaction with centrality and reciprocity, our results show that their effects are in the theorized direction and future studies may find these factors to be of importance.

While a number of studies have discussed the role of echo chamber type networks in radicalization, to date, no other study has included network-level factors at the individual level. The combining of social network analysis and survey data represents a convergence of social network analysis techniques that enable a deeper understanding of the objective and subjective factors pertaining to the online experience and online radicalization specifically.
Future studies examining the effects of personalization algorithms and the potential for algorithmic deviancy amplification should seek to examine larger samples. Additionally, stronger treatments to suppress algorithms, perhaps through the use of specially designed software may provide a better understanding of their direct and indirect effects on radicalization. While laboratory experiments may be able to better control algorithms or implement experimental platforms, the approach taken in the current study was one which preferred for allowing natural usage of a social media platform. Nevertheless, future studies may also seek to analyze network and belief development over longer observation periods as well.

11. References


Gilbert, E., Bergstrom, T., & Karahalios, K. (January 2009). Blogs are echo chambers: Blogs are echo chambers. *HICSS’09. 42nd Hawaii International Conference on IEEE* (pp. 1-10). System Sciences.


shape the evaluation of right-wing extremist internet propaganda. *Journal for deradicalization*, (10), 203-229.