Towards a Story Scheme Ontology of Terrorist MOs

J.G.T. Peters
Utrecht University & Netherlands National Police
The Netherlands
j.g.t.peters@uu.nl

F.J. Bex
Utrecht University & Tilburg University
The Netherlands
f.j.bex@uu.nl

***Abstract*—Crime investigation and criminal intelligence analysis often rely on the notion of *modus operandi*. We propose modelling such MOs as story schemes and show how real-life terrorist incidents can be assigned to such schemes. This is intended to support sensemaking activity performed by police analysts. We discuss several requirements of MO schemes and present an implementation in the form of an OWL ontology.**

I. Introduction

Modern policing paradigms increasingly emphasise the leading role of criminal intelligence analysis [19], a collection of sensemaking [18] tasks performed by practitioners like police analysts. Practitioners collect and document insights regarding criminal activity. Collaborative networks are mapped and internal conflicts are studied. Insights are used to prioritise investigations and indicate information gaps. Much focus is on prevention by hindering criminal activities, especially in the counter-terrorism domain.

Terrorism prevention and police investigation in the counter-terrorism domain require more than the insight that someone is a radical extremist. Even if we know someone recruits dangerous zealots specifically, there are many modes of involvement. Some methods might be typical of certain jihadist groups, whereas certain violent environmentalists or ultra-nationalists might operate in a wholly different fashion. Ways of conducting crimes are known as *modi operandi* (MOs) and recognising them is essential to law enforcement. Criminals change their MO as they develop. Certain regions or location types can be particularly facilitating of specific MOs. A strong similarity in MO between multiple criminal cases can be used to attribute them to the same individual or group ('crime/case linkage' or 'linkage analysis' [11]). Shifts in MO throughout criminal society may be noted and even anticipated through careful analysis. It is vital that MOs be correctly assigned to criminal subjects in such a way as to allow the study of variation and similarity of MOs, both between subjects and over time.

Recognising a crime as following a certain MO, thereby allowing for various sensemaking tasks, requires that the MO in question be already familiar to practitioners. Furthermore, the attribution of crimes and criminals to MOs by an entire intelligence organisation requires a consensus on those MOs. This implies the necessity of a body of knowledge we would call an 'MO knowledge base'. As new incidents occur that do not strictly follow known MOs, this knowledge base would have to incorporate these. Similarly, as evidential details about a case are revealed and cases are linked, practitioners might feel justified in specifying new, more detailed versions of existing MOs. In time, this growing body of knowledge would encourage further sensemaking support through data science techniques and the development of knowledge-based systems.

An MO knowledge base must therefore be developed to which real-life cases can be mapped by practitioners. An MO knowledge base requires a definition of what exactly constitutes an MO. The mapping of incidents to knowledge structures must be useable to practitioners. Practitioners are familiar with informal representations like mind maps and simple network graphs [22]. Any formal representation must closely match the common-sense understanding of practitioners. We represent our knowledge base in an OWL ontology [15], to which we assign real-life terrorist incidents. The incident data we use in our prototype ontology stems from the Global Terrorism Database, a publicly available[1] dataset with terrorism-related incident data [16]. The complete dataset contains 181,691 entries, described in 135 features.

The development of an MO knowledge base is conceptually related to other sensemaking developments we see in police practice, in particular the tendency towards standardisation of notions like criminal networks, logistical processes and illicit markets. This allows practitioners to combine theoretical, criminological insights with specific facts about known criminals in large-scale, integrated, high quality knowledge bases—known as 'intel positions' by practitioners. This then becomes the basis for further reasoning tasks. Examples of relevant reasoning tasks are the imputation of missing values (e.g. weapons used or likely gang affiliation), the consistent handling of missing or contrary evidence (e.g. witness testimonies or fingerprints) and analyses based on the completeness of stories (e.g. MOs or alibis).

The rest of this paper is structured as follows. We study the concepts related to MOs in Section II. Section III presents the beginnings of a knowledge base and briefly discusses attempts to support its construction through data mining techniques. In Section IV, we look at literature that utilises some of these concepts. Limitations of these demonstrative experiments and possible future directions of our research are described in Section V.

[1] GTD website: https://www.start.umd.edu/gtd/

II. *Modi Operandi*

From our experience, law enforcement practitioners typically follow one of two interpretations of MOs: an MO is formulated as either a sequence of states and/or actions or a set of circumstances such as locations, vehicles, weapons, victims. We will discuss both perspectives. Regardless, MOs typically describe the situational demands necessary to achieve the crime's goals and not those which are merely incidental. Symbolic or ritualistic acts may be excluded from the actual MO in certain criminological studies [11].

In practice, an MO's level of detail is typically restricted by the necessity to be able to usefully differentiate it from functionally distinct crimes and to be able to usefully link functionally identical crimes. It seems intuitive to state that something as broad as 'killing people with bombs' is insufficient to be an MO, as there is little value in attempting to link all bombings on an investigative level. Yet describing the most specific, evidently unique details of an actual incident as part of its MO prohibits any case linkage. The exact time and place or the backgrounds of victims may vary between incidents, for example, yet those details are likely to be deemed irrelevant for the MO.

A. *Sequences*

Stories organise facts into coherent hypotheses, and both story-based reasoning and evidence-based argumentation are common throughout forensic reasoning [2, 12, 17, 20]. Stories are constructed based on initial evidence and are subsequently compared, attempting to find the story that matches the evidence and meets criteria such as plausibility—"the extent to that it corresponds to the decision maker's knowledge about what typically happens in the world and does not contradict that knowledge" [17, p. 130].

The sequence-based understanding of MOs assumes that MOs are a particular type of story. There exist many perspectives on the nature of stories and how they are structured. Schank & Abelson [21] suggest that stories ('scripts') structure our knowledge about the world, by helping us fill in missing information. They describe the restaurant script, which contains knowledge structures for the procedures expected in a restaurant and can be further specified in a fast food track, a cafeteria track, and so on. These scripts are very similar to 'episode schemes' [17].

Figure 1 depicts a simplified version of Pennington and Hastie's episode scheme for intentional actions, which functions as a standard story structure. These can be seen as a kind of 'story schemes' [3]. This brings us to Definition 1—for a formal definition, see [2].

Definition 1. *[Story scheme] A story scheme is list of states and/or actions, described at a certain level of abstraction together with the possible causal relations between them, where each story scheme is associated with at least one story which is an instance of it.*

Story schemes can be modelled as lists of elements that range from the general to the specific, connected through causal links. Bex [3] requires that story schemes follow the simplified version of Pennington and Hastie's scheme for intentional actions and thereby correspond with Schank and Abelson's belief-goal-plan-action chain.

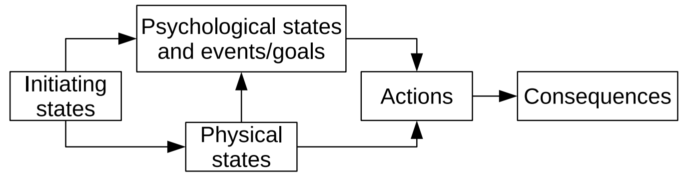

Figure 1. Pennington and Hastie's episode scheme for intentional actions. All arrows indicate causal relationships.

Figure 2 depicts an illustration of the mapping between an incident and the intentional actions scheme using abstractions. Abstractions are proxies for more elaborate inferences. For example, an explosion in public might lead us to infer that a bomb was detonated with the intention to harm people. That implies that this explosion was a deliberate bombing. This inference may be wrong. The detonation may have been an accident and we may have to view it in a different light.

Thinking of MOs as story schemes provides several advantages, including: a way to express whether a crime fits an MO (all abstractions are plausible), a way to express the plausibility of the MO (it fits the intentional actions scheme and all internal causal links are plausible), and a minimal requirement on the details of its causal structure.

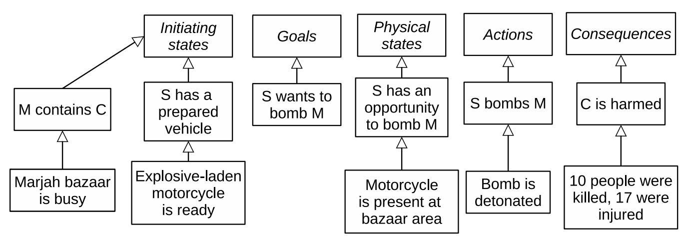

Figure 2. The mapping of an incident (bottom level) to an abstracted story scheme (middle level) and its mapping to an episode of the intentional actions scheme (top level). Abstractions are represented by white arrow heads; causal links have been left implicit. This describes how the set of suspects S use a vehicle to bomb marketplace M in order to harm a set of civilians C.

Viewing MOs as story schemes is similar to the criminological notion of 'crime scripts' proposed by Cornish [8]. Crime scripts are stepwise processes that describe how a crime is committed. Crime scripts are an adaptation of Schank & Abelson's cognitive scripts [21]. Like their tracks for the restaurant script, a robbery from a person is a form of robbery, and that script can be further specified as a subway mugging (see Table I) [8]. Therefore, crime scripts exist within a taxonomical structure.

A crime script contains several process steps known as 'scenes'. The number of scenes is not fixed and the order in which they occur is not necessarily fixed, as it may contain contingencies [8]. Scripts can be nested, since each scene can be a simpler script of its own. Thus scripts are contained

Table I
THE 'SUBWAY MUGGING' TRACK, AFTER [8].

Protoscript	Robbery
Script	Robbery from person
Track	Subway mugging
Script scene	**Script action**
Preparation	Meet and agree on hunting ground
Entry	Entry into underground system
Pre-condition	Travel to hunting ground
Pre-condition	Waiting/circulating at hunting ground
Instrumental pre-condition	Selecting victim and circumstance
Instrumental initiation	Closing-in/preparation
Instrumental actualisation	Stringing at victim
Instrumental actualisation	Pressing home attack
Doing	Take money, jewellery, etc.
Post-condition	Escape from scene
Exit	Exit from system

within other scripts through aggregation as well as parent-child relationships. Crime scripts not only express the procedural aspects of a crime; they may also express requirements like actions, casts, props and locations.

Crime scripts have been used in several criminological studies since their conception [1, 14]. Cornish's [8] proposed use of paths, levels and permutations are not as widely adopted as the idea of scripts itself and the connection to Schank & Abelson's work [21] can be tenuous in practice.

B. Circumstances

As police analysts typically think of an MO as a summarised narrative, it is frequently presumed that it should contain a minimum number of descriptors known as 'elements of circumstance' or just 'circumstances'. These are traditionally expressed as sets of interrogative words, ranging from three (e.g. *who, what, where*) to the classical seven circumstances (adding *when, why, in what way* and *by what means*), and beyond.

Thinking of MOs as circumstances comes with its own advantages. Law enforcement agencies often structure their knowledge of crimes and criminals according to some conceptual model, which allows for the study of phenomena or of social networks (e.g. [9]). Circumstances naturally form elements of interest to such analyses. They allow practitioners to notice connections between crimes and location types, or groups and specific means such as vehicle types.

When practitioners use the classical set, the distinction between *what* (description in sum), *in what way* (behavioural description) and *by what means* (instrumental description) can occasionally become superfluous. The obvious dichotomy of perpetrator and target/victim often seems to warrant a distinction between subject (*who*) and object (*whom*).

'Market squares during peak hours' seems an appropriate level of detail for *where* and *when*, as opposed to specifying an actual market and time of day. Intuitively, 'public places during daytime' would be a more generic version of this MO, describing a greater variety of incidents. Changing the scope of a circumstance, as in referring to the country instead of the market square, seems undesirable. MOs such as car bombings may be common in some country and there may be encouraging conditions there, but the MO in no way functionally requires that country.

The interpretations of MOs as either circumstances or story schemes are not mutually exclusive. Story schemes, including (crime) scripts [8, 21], do sometimes refer to circumstantial requirements such as locations and items. The combination of the two perspectives into one notion of story schemes for MOs is quite straightforward.

Definition 2. *[MO scheme] An MO scheme is a named story scheme (Definition 1) depicting an MO as a series of elements, where each element:*

- *Describes a state/action*
- *Describes all circumstances which are relevant to that state/action, where possible circumstances are subtypes of:*
  - *Actor (perpetrator or target/victim)*
  - *Location*
  - *Time*
  - *Means*
  - *Motive*

As suggested by Definition 2, the actions of an MO scheme mention circumstances as attributes, to just those levels of abstraction and detail which matter to the MO. In this scheme, *what* refers only to the scheme's name and *in what way* is described entirely by series of states/actions—see Figure 3.

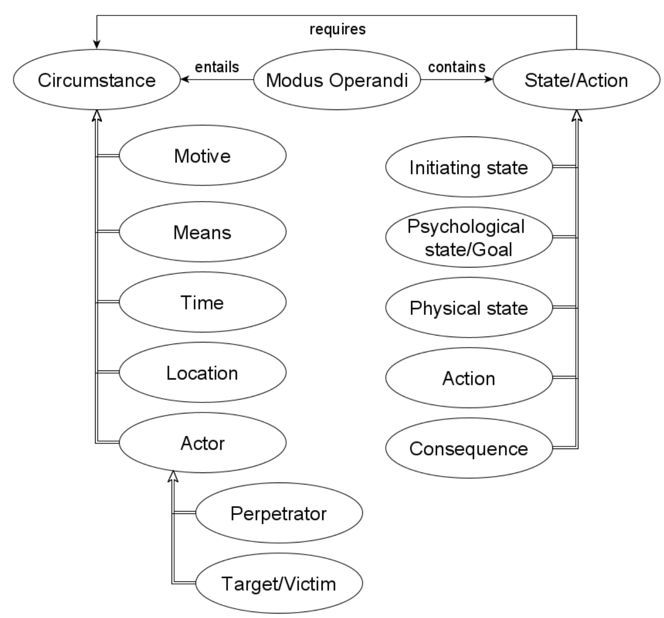

Figure 3. The top of the ontological model implied by Definition 2, including object properties (black arrow) and class-subclass relationships (white arrows).

III. MO KNOWLEDGE BASE

Our envisioned MO knowledge base contains a diverging hierarchy of MOs. This is essentially a taxonomy of the MOs used by criminals. A sensemaking tool, through which practitioners can map incidents to MOs and specify new ones, would facilitate the growth of this taxonomical knowledge base. We define this taxonomy and its internal relationships as follows.

Definition 3. *[Abstraction of MO schemes] An MO scheme (Definition 2) is one of n subclasses (children) of a parent class, where:*

- *Each action of the parent scheme is either equal to or an abstraction of one or several actions of the child scheme*
- *Each circumstance belonging to the actions of the parent scheme is either equal to or an abstraction of the circumstances of the child's actions*

Definition 4. *[MO taxonomy] For a given criminal domain, an MO taxonomy is the collection of MO schemes (Definition 2) connected through abstractions (Definition 3) which represents the body of knowledge concerning known MOs of that domain.*

Thanks to the ever-growing knowledge base of MOs in accordance with Definition 4, centrally accessible to an intelligence organisation, criminological matters of interest would naturally be produced, e.g. the observation that certain types of criminal tend to professionalise and converge on very specific MOs. When MOs are not treated in a similar manner and are instead entities that do not belong to any centralised knowledge base, a valuable source of insights is ignored.

We have constructed a prototype ontology in OWL [15] in accordance with the definitions provided in the current paper. This can be viewed as a step towards moving beyond sensemaking and towards a knowledge-based system, where the system itself is capable of reasoning tasks. Ontology reasoners allow for the automatic classification of both classes and instances ('individuals' in OWL terminology), and the identification of inconsistencies.

The top-level concepts of our ontology correspond to Figure 3 and the states/actions are related in accordance with Figure 1. This creates a taxonomy of MOs that follows the intentional actions scheme. The full ontology and instructions how to visualise it are available as an appendix[2].

Our ontology includes a subset of GTD instances. MO instances are added to the ontology based on features which correspond to elements of circumstance, which are themselves also added as instances or classes within the ontology. States/actions cannot be automatically derived from the GTD and require a practitioner's interpretation. The incident's summary is used as the description. Figure 4 shows a visualisation of our ontology containing a mapping of GTD instances.

A. Handling Unknowns

The GTD's incidents contain unknowns, i.e. missing variables. This is to be expected, as not every feature can be accurately determined for every incident. Some incidents are never claimed by a terrorist organisation, exact perpetrator numbers may be difficult to verify and the precise weapons used can remain unclear. This is representative of the incident data that

[2]Ontology: https://bitbucket.org/JGTP/mo_gtd

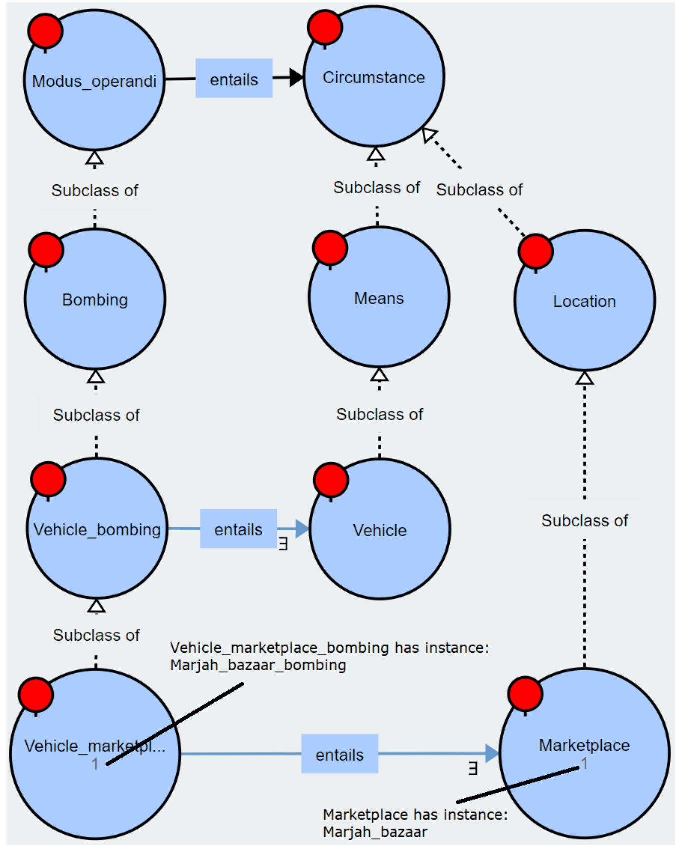

Figure 4. A screenshot of a small subset of our ontology. This illustrates how the MO of Figure 1 can be viewed in light of our proposed knowledge base.

would be handled by police practitioners. We therefore have to consider how missing data relates to our proposed MO knowledge base.

In practice, law enforcement would find ways of dealing with such uncertainties. Missing data can be imputed from the available data for the incident within a certain degree of accuracy. When it is a categorical feature that is missing, imputing from other data features it is a straightforward classification task. For instance, this makes it possible to predict the weapon used from other variables (as demonstrated by De Kock [13]). When we treat the perpetrator group as the target variable, it may be possible to identify the responsible terrorist organisation through their choice of weapon, victim, etc.

For instance, as part of our data exploration of the GTD, we have trained a random forest [5] classifier (100 estimators) using three-fold cross-validation on the distinction between the Taliban ($n = 8723$) and the Islamic State of Iraq and the Levant (ISIL) ($n = 6385$) using only those GTD features we would use as circumstances or states/actions in an MO knowledge base[2]. We avoid the usage of trivial indicators of the responsible terrorist organisation, such as the country where incidents took place. We also prevent the classifier from using the incident's year and restrict the dataset to the post-1997 GTD data collection effort, to keep the availability of features consistent. We find this task performs reasonably well, averaging at a weighted F1 score of 0.773 and an ROC-AUC of 0.843. Feature importance scores suggest a difference in the number of casualties (the Taliban seem to make fewer but suffer more than ISIL), specific weapons (the Taliban use more landmines, ISIL more vehicle bombs) and target types (the Taliban target more police and fewer citizens than ISIL) and the reporting of multiple related incidents (slightly more common for ISIL). This could hint at an inherent difference in MO between the two organisations, perhaps showing how the two groups are engaged in a subtly different type of power struggle. This exploratory classification illustrates how data mining could provide ways of handling unknowns and complementing knowledge of terrorist MOs.

The classification of MOs can similarly be treated as a data mining task. However, Definition 2 implies that labelling instance can be strictly rule-based—our mapping of GTD incidents to MOs is simply based on the values of the available circumstances for each incident. It therefore makes more sense to see machine learning classification as complementary to a rule-based approach. This approach is best achieved by producing an ontology of MOs and assigning incidents to it, which can be (partially) automated. Not only would an ontology reasoner be able to classify incidents in such a way as to suggest additions to the knowledge base; it is possible to produce a Bayesian network from such an ontology [10], thereby generating a classifier that can handle uncertainties.

IV. Related Work

De Kock [13] constructs a dataset through feature engineering from the GTD. His new set is a collection of descriptions of the historical records using twelve film industry-inspired ‘scenario elements’. These abstract elements are sets of more concrete features. De Kock’s goal is to create a ‘scenario model’ by which to anticipate criminal behaviour, but he relates neither his model nor his element called *modus operandi* to crime scripts. Some of these elements are rarely known during terrorist attacks. However, De Kock states that the few which might be known can help impute the likely values of the others. For this straightforward classification task, he uses a decision tree approach, selected based on exploratory results.

Recognising MOs from incidents could be supported through NLP techniques. For example, the recognition of an MO could be performed through schema induction. Chambers & Jurafsky present an unsupervised method of learning narrative event chains, which they define as chains with a single protagonist [6]. In later papers, they show that it is possible to learn schemas from texts without predefining them [7].

Academic studies of terrorist MOs using crime scripts as well as actual police data have been performed. De Bie et al. [4] take police investigation reports on 51 ‘foreign fighters’ (radical jihadists who try to mobilise themselves to join a foreign conflict) and study variations in their execution of their script. They find that these mobilisation attempts can be grouped into three ‘episodes’. Per crime script scene, they then compare these three groups and study additional differences between them. For example, the invasion of Iraq and Afghanistan changed the orientation stage of the crime script, in that Islam was now perceived as being under explicit military threat by the West. In the earlier episode, subjects were instead concerned with their religious-moral perception of supposedly Western-influenced rulers of Muslim countries. Overall, they conclude that the situational factors of geopolitical situation, social opportunity structures and technological developments are key to the variation they observe.

V. Discussion and Future Work

The MO knowledge proposed in this article can be related to similar developments we see in practice at the police. Our ontology is only an exploratory study of what an MO ontology could look like in OWL. This is knowledge representation, but we would be interested in developing this into an analysis process. In future, the ontology could grow much larger and more elaborate, especially if various practitioners were to make use of sensemaking software tied to the ontology. We would be interested in studying its uses as a knowledge base. In the process, we could endeavour to map the entire GTD to the ontology.

We would be particularly interested in studying the reasoning tasks that could be performed when practitioners have access to a large-scale ‘intel position’, in which theoretical insights and case details are combined. A requirement from practice which we have only briefly mentioned is that barriers be identifiable through which to obstruct the MO. Barriers play a central role in law enforcement policy, yet it is unclear how exactly these must be viewed in light of precisely defined MOs. In one possible direction of future work, we would like to take a formal, general approach in reasoning with such barriers and consider further refinements to our ontology which might be necessary to incorporate them.

Hierarchical clustering (e.g. [23]) is one possibility of supporting the sensemaking process whereby the MO knowledge base is expanded with new MOs. Such clustering groups instances recursively, based on their similarities. The result is a taxonomical structure (’dendrogram’). Under the right conditions, the result might be similar to a practitioner-formulated taxonomy, given the same set of incidents.

Finally, the circumstances used in our MOs can each be further subdivided into more specific classes, such as various roles for actors involved in the MO. This allows us to construct complex MOs, which is necessary for modelling complex crimes, such as the large-scale laundering of illegally obtained funds.

Conclusion

We have suggested that story schemes based on the intentional actions scheme may be suitable candidates for modelling MOs in criminal intelligence analysis and have listed several requirements on these schemes, including the incorporation of elements of circumstance. This results in a proposed knowledge base by which MO classification may be automated, which we used to construct an early OWL ontology supplemented with instances from the Global Terrorism Database.

References

[1] E. Beauregard, J. Proulx, K. Rossmo, B. Leclerc and J.-F. Allaire, 'Script Analysis of the Hunting Process of Serial Sex Offenders', *Criminal Justice and Behavior*, vol. 34, no. 8, pp. 1069–1084, 2007, ISSN: 0093-8548.

[2] F. J. Bex, *Arguments, stories and criminal evidence: a formal hybrid theory*, ser. Law and philosophy library v. 92. Dordrecht ; New York: Springer, 2011, OCLC: ocn663950184, ISBN: 978-94-007-0139-7.

[3] F. J. Bex, 'Analysing Stories Using Schemes', in *Legal Evidence and Proof*, London: Routledge, 2016, pp. 107–130.

[4] J. L. de Bie, C. J. de Poot and J. P. van der Leun, 'Shifting Modus Operandi of Jihadist Foreign Fighters From the Netherlands Between 2000 and 2013: A Crime Script Analysis', *Terrorism and Political Violence*, vol. 27, no. 3, pp. 416–440, 2015, ISSN: 0954-6553.

[5] L. Breiman, 'Random Forests', *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, ISSN: 1573-0565.

[6] N. Chambers and D. Jurafsky, 'Unsupervised Learning of Narrative Event Chains', in *Proceedings of ACL-08: HLT*, Columbus, Ohio: Association for Computational Linguistics, 2008, pp. 789–797.

[7] N. Chambers and D. Jurafsky, 'Template-Based Information Extraction without the Templates', in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA: Association for Computational Linguistics, 2011, pp. 976–986.

[8] D. Cornish, 'The procedural analysis of offending and its relevance for situational prevention', *Crime prevention studies*, vol. 3, pp. 151–196, 1994, ISSN: 1065-7029.

[9] P. A. C. Duijn, 'Detecting and disrupting criminal networks: A data driven approach', PhD, University of Amsterdam, 2016.

[10] S. Fenz, 'An ontology-based approach for constructing Bayesian networks', *Data & Knowledge Engineering*, vol. 73, pp. 73–88, 2012, ISSN: 0169-023X.

[11] R. R. Hazelwood and J. I. Warren, 'Linkage analysis: Modus operandi, ritual, and signature in serial sexual crime', *Aggression and Violent Behavior*, vol. 8, no. 6, pp. 587–598, 2003, ISSN: 13591789.

[12] R. J. Heuer, *Psychology of Intelligence Analysis*. Center for the Study of Intelligence (U.S.), 1999, ISBN: 0-16-059035-3.

[13] P. de Kock, 'Anticipating criminal behaviour: Using the narrative in crime-related data', PhD, Tilburg University, Tilburg, the Netherlands, 2014.

[14] B. Leclerc, 'Script Analysis for Crime Controllers: Extending the Reach of Situational Crime Prevention', in *Organized Crime, Corruption and Crime Prevention: Essays in Honor of Ernesto U. Savona*, S. Caneppele and F. Calderoni, Eds., Cham: Springer International Publishing, 2014, pp. 13–20, ISBN: 978-3-319-01839-3.

[15] D. L. McGuinness and F. van Harmelen, 'OWL Web Ontology Language Overview', Tech. Rep., 2004.

[16] National Consortium for the Study of Terrorism and Responses to Terrorism (START), *History of the GTD*, 2018.

[17] N. Pennington and R. Hastie, 'Reasoning in explanation-based decision making', *Cognition*, vol. 49, no. 1, pp. 123–163, 1993, ISSN: 0010-0277.

[18] P. Pirolli and S. Card, 'The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis', in *Proceedings of International Conference on Intelligence Analysis*, 2005.

[19] J. H. Ratcliffe, *Intelligence-Led Policing*. Routledge, 2016, ISBN: 978-1-317-51040-6.

[20] M. Saletta, A. Kruger, T. Primoratz, A. Barnett, T. van Gelder and R. E. Horn, 'The role of narrative in collaborative reasoning and intelligence analysis: A case study', *PLoS ONE*, vol. 15, no. 1, 2020, ISSN: 1932-6203.

[21] R. C. Schank and R. P. Abelson, 'Scripts, Plans, and Knowledge', in *Proceedings of the 4th International Joint Conference on Artificial Intelligence - Volume 1*, ser. IJCAI'75, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1975, pp. 151–157.

[22] N. Timmers, 'The hybrid theory in Practice: A case study at the Dutch police force', Master's thesis, Utrecht University, 2017.

[23] J. H. Ward, 'Hierarchical Grouping to Optimize an Objective Function', *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963, ISSN: 0162-1459.