CHEMICAL ENGINEERING TRANSACTIONS
VOL. 116, 2025
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Bruno Fabiano, Valerio Cozzani

ISBN 979-12-81206-18-2; **ISSN** 2283-9216
DOI: 10.3303/CET25116064

An Innovative Approach to the Identification of Reference Safety and Security Scenarios in the Process Industry

Matteo Iaiani*, Alessandro Tugnoli, Valerio Cozzani

LISES – Department of Civil, Chemical, Environmental, and Materials Engineering, Alma Mater Studiorum – University of Bologna, via Terracini n.28, 40131 Bologna (Italy)
matteo.iaiani@unibo.it

Hazard identification is a critical step in both safety and security risk assessments. Traditional approaches predominantly rely on historical data and Exploratory Data Analysis (EDA) to define reference scenarios. However, EDA lacks standardized methods for identifying and ranking incident chains. To overcome this limitation, the BAS^{2}E (Bayesian network Analysis of Safety and Security Events) methodology was developed to systematically derive reference scenarios from historical event data using Bayesian Networks (BN). The methodology uses statistical data from accident records and applies the Noisy-OR gate model to manage uncertainties in the specification of conditional probability tables (CPTs). Sensitivity analysis is used to quantify the influence between nodes in the BN, enabling a ranked selection of the most critical incident chains for inclusion in risk assessment. The methodology is demonstrated using a dataset of 109 security incidents occurred in the offshore Oil&Gas industry.

1. Introduction

The storage and processing of large quantities of hazardous materials in chemical, process, and Oil&Gas industries (offshore and onshore) can lead to significant incidents such as releases, explosions, and fires, resulting in severe consequences for human life, the environment, and the assets. These events can be triggered by internal system-related factors (safety) (Mannan, 2012) or intentional attacks (security) (Iaiani et al., 2024), and their proper identification and quantitative assessment is of paramount importance to enhance safety and security and to prevent potential disasters.

A common high-level approach to gather valuable insights supporting existing qualitative and quantitative risk analysis procedures (e.g., safety QRA studies, security vulnerability/risk assessment (SVA/SRA) methodologies) is the analysis of past events that occurred in similar facilities (e.g., belonging to the same industrial sector) (Iaiani et al., 2023). In fact, this analysis can provide reference scenarios (chain of events from the origin of the risk to the final outcomes suffered by the affected facilities) that can be used by practitioners and authorities as the basis to undertake case-specific assessments. The detailed examination of past events plays a crucial role in understanding the multifaceted nature of these events, identifying, e.g., common causes, systemic vulnerabilities of the systems, and potential consequences.

Exploratory Data Analysis (EDA) is conventionally employed in past event analysis; however, it shows limitations in systematically analysing incident datasets and prioritizing relevant incidental chains of events (Sales et al., 2007). Specifically, traditional EDA approaches may struggle to capture the complex interdependencies between various factors contributing to safety and security incidents, not allowing to properly represent incident causation and evolution. Moreover, EDA often faces challenges related to data availability, quality, and consistency (Konstandinidou et al., 2011).

To overcome these limitations, the present study proposes a Bayesian Network modelling-based methodology, alternative to canonical EDA, for the systematic identification of the most relevant incidental chains, that serves as reference scenarios, from safety/security incident datasets. This methodology, called BAS^{2}E (Bayesian network Analysis of Safety and Security Events) leverages probabilistic models to explicitly represent and quantify the uncertainties and causal relationships between different variables involved in incidents. The use of

Paper Received: 16 November 2024; Revised: 9 February 2025; Accepted: 11 May 2025
Please cite this article as: Iaiani M., Tugnoli A., Cozzani V., 2025, An Innovative Approach to the Identification of Reference Safety and Security Scenarios in the Process Industry, Chemical Engineering Transactions, 116, 379-384 DOI:10.3303/CET25116064

the Bayesian Network modelling allows for the integration of both quantitative data and qualitative expert judgments, enhancing the robustness and relevance of the analysis in the face of data limitations. The proposed innovative approach is illustratively applied to a case study concerning a dataset of security-related incidents occurred in offshore oil and gas fluid production facilities, providing reference security scenarios to be used in the context of security studies of offshore critical infrastructures.

2. Mathematical models

2.1 Bayesian Network

Bayesian Networks (BNs) are probabilistic models that represent uncertain relationships among variables using a directed acyclic graph (DAG). Each node represents a variable, and directed edges between nodes signify conditional dependencies. The strength of these dependencies is quantified through Conditional Probability Tables (CPTs). The joint probability distribution of the set of variables $U = \{F_1, \dots, F_n\}$ in a BN is given by (Jensen and Nielsen, 2007):

$$Pr(U) = \prod_{i=1}^{n} p(F_i|Pa(F_i)) \tag{1}$$

where $Pa(F_i)$ represents the parent nodes of the variable F_i.
BN modelling uses Bayes' theorem to update probabilities dynamically when new evidence becomes available, providing the posterior probabilities as (Jensen and Nielsen, 2007):

$$Pr(U|E) = \frac{Pr(U \wedge E)}{Pr(E)} = \frac{Pr(U \wedge E)}{\sum_U Pr(U \wedge E)} \tag{2}$$

This dynamic updating makes BNs a powerful tool for accident modelling and risk assessments. Compared to traditional methods like Bow-Tie (BT) analysis, BNs are better suited for handling conditional dependencies and dynamically updating risk probabilities as system conditions evolve. For this reason, the BN modelling is used in the BAS^{2}E methodology as core mathematical model.

2.2 Noisy-OR gate model

A significant challenge in BN modelling is determining the CPTs, especially when dealing with limited data availability. The Noisy-OR gate model offers a solution by simplifying the CPT through reduced parameter requirements. In this model, each parent node X_i has a known probability p_i of causing a child node Y to be in state "YES," independently of other parent nodes. This reduces the number of parameters from 2^n to n for n parent nodes. The probability that Y is in state "YES" when several parent nodes are "YES" is given by (Oniśko et al., 2001):

$$\Pr\left(Y = YES \middle| \bar{X}_p\right) = 1 - \prod_{i:X_i \in \bar{X}_p} (1 - p_i) \tag{3}$$

In the present study, the Noisy-OR gate model is adopted within BAS^{2}E methodology as the mathematical model to calculate the conditional probabilities referred to multiple parent nodes in state "YES" under specific conditions (see Section 3.2 in the following). This model reduces the number of parameters required to define conditional probabilities, making it feasible to handle cases where empirical data is limited, while still ensuring that the probabilistic relationships in the system are adequately represented.

2.3 Sensitivity Analysis

Sensitivity analysis in Bayesian Networks identifies which variables most influence the probability of a target node being in a specific state. This is done by analysing how changes in the conditional probabilities of parent-child nodes affect the "probability of interest" of the target node. Mathematically, this probability is expressed as $\Pr(f_i|e)$, where f_i represents the state of the target node F_i, and e denotes the evidence observed in the network. The sensitivity of the probability of interest $\Pr(f_i|e)$ to changes in the conditional probability $x_{j,z} = p(f_j|f_z)$ (where f_j is a state of node F_j and f_z a state of node F_z which is parent of F_j) is measured by the derivative (Kjærulff and van der Gaag, 2000):

$$D_{i,j,z} = \frac{d}{dx_{j,z}}\left(\Pr(f_i|e)\right) \tag{4}$$

High values $D_{i,j,z}$ indicate a strong influence, while low values suggest a weaker influence. A zero value for $D_{i,j,z}$ implies no influence.

Based on this formulation, sensitivity analysis is used in the BAS²E methodology to identify the most relevant incident chains in the system.

3. The BAS²E methodology

3.1 General information

The BAS²E (Bayesian network Analysis of Safety and Security Events) methodology presented in this study is a structured four-step process aimed at identifying the most significant incident chains (the incidental scenarios of concern) using Bayesian Network analysis of past event datasets. These identified chains can be used as reference scenarios in both safety and security risk assessments (see Figure 1). In particular, in the safety domain (ref. to ISO 31000 framework), support can be provided to the "Risk Identification" phase, while in the security domain (ref. to API Recommended Practice 7807) to the "Threat Identification & Assessment" and "Vulnerability Identification & Assessment" phases.
In the next paragraph, each step of the BAS²E methodology is detailed.

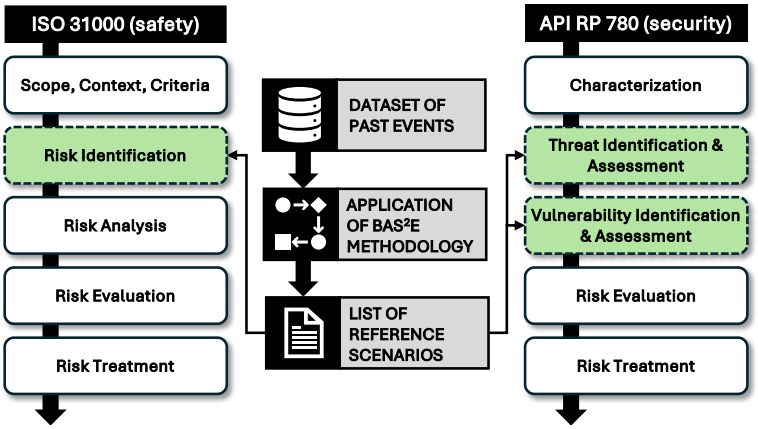

Figure 1: Integration of the BAS²E methodology with established frameworks for safety (ISO 31000) and security (API RP 780) risk assessments.

3.2 Detailed description

The flowchart of the BAS²E methodology is reported in Figure 2-a.
The first step involves creating the directed acyclic graph (DAG) that models the structure of the incident chain of interest. Figure 2-b shows a generic DAG related to a 3-layer incident chain (e.g., faults, release scenarios and consequences in the safety domain, or threats, attacks, and physical damage scenarios in the security domain). The determination of the type and number of layers within the chain structure depends on the quality of information present in the dataset entries for the specific case under analysis. The information for each layer category is organized into finite classes (e.g., in Figure 2-b, layer 1 has 3 classes that could be "random failure", "human error", "natural hazard" for the layer "faults" in the safety domain). In cases where data is incomplete, an "unknown" class is included. These classes are represented in the DAG as binary nodes (states "YES" and "NO") if they are not mutually exclusive, or as a single node with the classes as states if they are mutually exclusive. The causal relationships between nodes of two subsequent layers are depicted using arcs (see Figure 2-b).
The second step involves the specification of the conditional probability tables (CPT) for each node. For nodes without parent nodes (i.e., nodes of layer 1), the marginal probabilities are calculated as:

$$p_{CL_{1,j}} = \frac{\text{n° events categorized as } CL_{1,j}}{\sum_j \text{n° events categorized as } CL_{1,j}} \tag{5}$$

where j indexes the classes of layer $i = 1$.
For nodes with a single parent or with multiple parents but only one in state "YES", the conditional probabilities are determined as:

$$p_{CL_{i,j},CL_{i-1,z}} = \frac{\text{n° events categorized as } \left(CL_{i,j}\right) \wedge \left(CL_{i-1,z}\right)}{\text{n° events categorized as } CL_{i-1,z}} \tag{6}$$

where j indexes the classes of i^{th} layer, and z the classes of $(i-1)^{th}$ layer.

For nodes with multiple parents in state “YES”, conditional probabilities are calculated using either equation (7) or the Noisy-OR gate model (refer to equation (3) above), depending on the results of a statistical significance test (test of proportion (Lane et al., 2003)). If the test provides sufficient evidence to reject the independence hypothesis (null hypothesis), the dataset information is utilized in equation (7) to determine the conditional probabilities:

$$p_{CL_{i,j},(CL_{i-1,z})\wedge\ldots\wedge(CL_{i-1,k\neq z})} = \frac{\text{n° events categorized as } \left(CL_{i,j}\right) \wedge \left[\left(CL_{i-1,z}\right) \wedge \ldots \wedge \left(CL_{i-1,k\neq z}\right)\right]}{\text{n° events categorized as } \left(CL_{i-1,z}\right) \wedge \ldots \wedge \left(CL_{i-1,k\neq z}\right)} \tag{7}$$

where j indexes the classes of i^{th} layer, z and k ($k \neq z$) the classes of $(i-1)^{th}$ layer.
Otherwise, equation (3) is applied. By this approach, a more accurate probability estimation is achieved, avoiding unreliable zero values in CPTs, and thus addressing the critical issue in EDA application related to data availability for multiple events occurring simultaneously.
The third step involves performing BN sensitivity analysis (see Section 2.3). Specifically, the analysis examines combinations of three nodes (triplets) at a time from sequential layers. The target node is the one in the highest layer of the triplet, while the other two nodes are from the immediately lower levels. With reference to the generic DAG shown in Figure 2-b, since the incident chain is formed by 3 layers, a single $D_{i,j,z}$ is calculated for each node of layer 3 set as target:

$$D_{CL_{3,j},CL_{2,z},CL_{1,t}} = \frac{d\left(\Pr\left(CL_{3,j}\middle|e\right)\right)}{d\left(p\left(CL_{2,z}\middle|CL_{1,t}\right)\right)}, \forall j = 1,2,3;\ \forall z = 1,2;\ \forall t = 1,2,3 \tag{8}$$

These derivatives rank the most significant dependencies in the DAG, helping to pinpoint the most relevant incidents chains.
The final step (step 4) involves selecting incident chains based on the calculated derivatives and a defined threshold (TS, e.g., TS=0.01). This threshold acts as a cutoff, as incident chains that meet or exceed the threshold are considered relevant and the other are discarded. The selected chains are then arranged into tables, which collectively form the final output of the methodology.

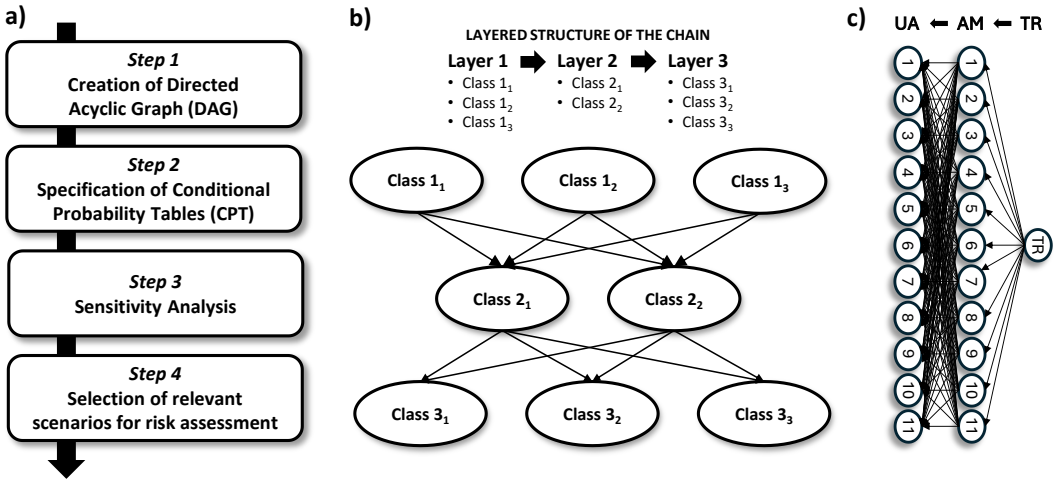

Figure 2: a) Flowchart of BAS²E methodology; b) Generic DAG referred to 4-layer incident chain; c) DAG obtained in Case Study (step 1). TR-codes, AM-codes, and UE-codes are defined in Section 4.2.

4. Case Study

4.1 Dateset definition

This case study shows the application of the BAS²E methodology to a dataset of 109 security incidents occurred in offshore Oil&Gas production facilities, all caused by intentional malicious attacks. The primary aim is to identify a set of reference security scenarios to be used in the context of security risk assessment of offshore Oil&Gas installations (e.g., the workflow proposed by API RP 70/70I (API, 2010).
The dataset, developed from a broad set of open databanks and scientific articles, is provided in a summarized version in the supplementary material of Iaiani et al. (2025).

4.2 Application of BAS²E methodology

The first step of the BAS²E methodology involves the definition of the chain structure and the development of the DAG. Three layers were considered, named “Threat types”, “Attack methods”, and “Undesirable events”. Information in the dataset was arranged to allow classification of the incidents into classes for each layer. In particular, the following classes were considered for “Threat types”: Terrorism/Guerrilla (TR1), Operations of military/paramilitary organization (TR2), Civil protest (TR3), Insurrection/Piracy (TR4), Cyber criminality (TR5), Insider threat (TR6), Unknown (TR7). Similarly, the classes considered for “Attack methods” are: Aircraft impact (AM1), Arson using incendiary means (AM2), Deliberate interference w/o aids (AM3), Shooting 1 (AM4), Shooting 2 (AM5), Use of explosives (AM6), Use of vessel bomb (AM7), Vessel impact (AM8), Unauthorized access (AM9), Robbery/Kidnapping (AM10), Cyber-attack (AM11), Unknown (AM12). Finally, those considered for “Undesirable events” are: Release (UE1), Explosion (UE2), Fire (UE3), Shootout (UE4), Operation shutdown (UE5), Pacific occupation (UE6), Seizing of workers (UE7), Near miss (UE8), Stolen goods (UE9), Loss of process (UE10), Other (UE11), Unknown (UE12). The developed DAG is shown in Figure 2-c: since the security threats are mutually exclusive, they have been represented as a single node (node “TR” in the DAG) with the classes as states, while AM-classes and UE-classes have been represented as single binary nodes (states “YES” and “NO”).

The second step of the BAS²E methodology involves the specification of the conditional probability table (CPT) for each node. In particular, CPT of TR-node (see DAG in Figure 2-c) was specified by applying eq. (5) using incident count from dataset. For all the other CPTs, conditional probabilities referred to a single parent in state “YES” were calculated by applying eq. (6) using dataset data, while in case of multiple parents in state “YES”, whenever a non-zero number of events was collected in the dataset, the test of proportion was applied and, based on the outcome, the Noisy-OR gate model (eq. (3)) or eq. (7) were adopted (an example of this approach is provided in Table 1 for the UE-node “Explosion (UE2)”). For all the other cases, the Noisy-OR gate model was directly applied to estimate conditional probabilities.

Table 1: Example of the approach adopted for quantification of a sub matrix of CPT of node “Explosion (UE2)”.

AM in state YES	Observed UE2 events given AM combinations	Expected UE2 events given AM combinations	p-value	Model to use	Calculated conditional probability
Shooting 2 (AM5) only	2	N.A.	N.A.	Eq. (6)	0.222
Use of explosives (AM6) only	12	N.A.	N.A.	Eq. (6)	0.923
Both AM5 and AM6	2	2.55	0.36 < α=0.9	Eq. (3)	0.940

The third step of the BAS²E methodology involves the application of sensitivity analysis as described in Section 2.3. Since the incident chain structure of interest has 3 layers, a single D_{UE_i,AM_j,TR_z} is calculated for each possible incident chain, taking the UE-nodes in their state “YES” as targets. For illustrative purposes, in Table 2 are reported the results for UE-node “Fire (UE3)” set as target.

The fourth step of the BAS²E methodology involves the selection of the most relevant incident chains through a cut-off criterion based on an absolute threshold value (TS) for the derivative D_{UE_i,AM_j,TR_z}. In the present study, TS was assumed to be equal to the minimum change in the proportion among the observed data, i.e. 1/(109-1)=0.009. For example, with reference to Table 2 (see the fifth column), the chains “Terrorism/Guerrilla (TR1)→”Arson using incendiary means (AM2)”→”Fire (UE3)” and “Operation of military/paramilitary organization (TR2)→”Shooting 2 (AM5)”→”Fire (UE3)” were selected as relevant (D equal to 0.42 and 0.08 respectively). On the contrary, the chains “Operation of military/paramilitary organization (TR2)→”Use of explosives (AM6)”→”Fire (UE3)” and “Terrorism/Guerrilla (TR1)→”Use of vessel bomb (AM7)”→”Fire (UE3)” were not selected as the respective derivative values is lower than the defined TS (D equal to 0.0029 and $4.76 \cdot 10^{-18}$ respectively).

Overall, the analysis reveals that the most critical attack methods (potential to cause severe events like explosions and fires), such as the use of explosives, use of vessel bombs, and arson, are part of relevant incident chains of “Terrorism/Guerrilla (TR1)” and “Operation of military/paramilitary organization (TR2)” security threats, which thus pose significant threat to offshore facility that could be subjected by these adversaries. Conversely, civil protest events resulted less critical, as all the relevant incident chains identified end with events having consequences of limited severity. In particular, the most probable attack method for this threat is “Unauthorized access (AM9)” followed by “Occupation (UE6)” of the asset. All the identified incident chains can be adopted as reference scenarios in the context of a SVA study for an offshore Oil&Gas fluid production facility. Moreover, this application proved the ability of the BAS²E methodology in overcoming most of the EDA limitations outlined above, specifically as regards data availability, chain prioritization, and systematization.

Table 2: Derivatives $D_{i,j,z}$ calculated in sensitivity analysis for node "Fire (UE3)" set as target.

Threat type (EL_1)	Attack method (EL_2)	Undesirable event (EL_3)	Calculated $D_{i,j,z}$	Selected
Terrorism/Guerrilla (TR1)	Arson using incendiary means (AM2)	Fire (UE3)	0.42	Yes
Terrorism/Guerrilla (TR1)	Shooting 2 (AM5)	Fire (UE3)	0.32	Yes
Insurrection/Piracy (TR4)	Arson using incendiary means (AM2)	Fire (UE3)	0.32	Yes
Operation of military/paramilitary organization (TR2)	Shooting 2 (AM5)	Fire (UE3)	0.08	Yes
Operation of military/paramilitary organization (TR2)	Use of explosives (AM6)	Fire (UE3)	0.0029	No
Terrorism/Guerrilla (TR1)	Use of vessel bomb (AM7)	Fire (UE3)	$4.76 \cdot 10^{-18}$	No

5. Conclusions

The BAS^{2}E methodology enables the derivation of reference scenarios from the Bayesian Network (BN) analysis of historical safety or security event datasets that allow for a structured classification of the sequence of events occurring during incidents. The Noisy-OR gate model is proposed in BN quantification to effectively address data limitations when observed data are not statistically significant, which is common in small datasets. The application of the methodology to a case study involving 109 security-related incidents occurred in the offshore Oil&Gas sector, demonstrated its capability to overcome the limitations of traditional statistical analyses such as Exploratory Data Analysis (EDA), which typically rely on filtering and counting techniques. The use of a derivative-based parameter in chain ranking was demonstrated to effectively capture changes in causal relationships between variables in the BN, offering a robust criterion for chain prioritization and selection. The identified incident chains can serve as reference scenarios in SVA studies (e.g., API RP 70) for offshore production platforms.

Acknowledgments

This work was supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union – NextGenerationEU.

References

American Petroleum Institute (API), 2010, API RP 70: Security for Offshore Oil and Natural Gas Operations.

Iaiani M., Fazari G., Tugnoli A., Cozzani V., 2025, Identification of reference security scenarios from past event datasets by Bayesian Network analysis, Reliability Engineering & System Safety, 254, 110615.

Iaiani M., Sorichetti R., Tugnoli A., Cozzani V., 2024, Modelling standoff distances to prevent escalation in shooting attacks to tanks storing hazardous materials, Reliability Engineering & System Safety, 241, 109689.

Iaiani M., Tugnoli A., Cozzani V., 2023, Identification of cyber-risks for the control and safety instrumented systems: a synergic framework for the process industry, Process Safety and Environmental Protection, 172, 69–82.

Jensen F.V., Nielsen T.D., 2007, Bayesian networks and decision graphs, 2nd ed., Springer, New York.

Kjærulff U., van der Gaag L.C., 2000, Making Sensitivity Analysis Computationally Efficient, in: Sixteenth Annual Conference on Uncertainty in Artificial Intelligence, 317–325.

Konstandinidou M., Nivolianitou Z., Kefalogianni E., Caroni C., 2011, In-depth analysis of the causal factors of incidents reported in the Greek petrochemical industry. Reliability Engineering and System Safety, 96, 1448–1455.

Lane D.M., Scott D., Hebl M., Guerra R., Osherson D., Zimmer H., 2003, Introduction to Statistics, Rice University.

Mannan S., 2012, Lees' Loss Prevention in the Process Industries: Hazard Identification, Assessment and Control, 4th ed., Elsevier, UK: Butterworth-Heinemann.

Oniśko A., Druzdzel M.J., Wasyluk H., 2001, Learning Bayesian network parameters from small data sets: application of Noisy-OR gates, International Journal of Approximate Reasoning, 27, 165–182.

Sales J., Mushtaq F., Christou M.D., Nomen R., 2007, Study of major accidents involving chemical reactive substances analysis and lessons learned, Process Safety and Environmental Protection, 85, 117–124.