
JOURNAL OF APPLIED LOGICS - IFCoLoG
JOURNAL OF LOGICS AND THEIR APPLICATIONS

Volume 9, Number 4

July 2022

Disclaimer

Statements of fact and opinion in the articles in Journal of Applied Logics - IFCoLog Journal of Logics and their Applications (JALs-FLAP) are those of the respective authors and contributors and not of the JALs-FLAP. Neither College Publications nor the JALs-FLAP make any representation, express or implied, in respect of the accuracy of the material in this journal and cannot accept any legal responsibility or liability for any errors or omissions that may be made. The reader should make his/her own evaluation as to the appropriateness or otherwise of any experimental technique described.

© Individual authors and College Publications 2022
All rights reserved.

ISBN 978-1-84890-386-9

ISSN (E) 2631-9829

ISSN (P) 2631-9810

College Publications
Scientific Director: Dov Gabbay
Managing Director: Jane Spurr

<http://www.collegepublications.co.uk>

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise without prior permission, in writing, from the publisher.

EDITORIAL BOARD

Editors-in-Chief
Dov M. Gabbay and Jörg Siekmann

Marcello D'Agostino	Melvin Fitting	Henri Prade
Natasha Alechina	Michael Gabbay	David Pym
Sandra Alves	Murdoch Gabbay	Ruy de Queiroz
Arnon Avron	Thomas F. Gordon	Ram Ramanujam
Jan Broersen	Wesley H. Holliday	Chrtian Retoré
Martin Caminada	Sara Kalvala	Ulrike Sattler
Balder ten Cate	Shalom Lappin	Jörg Siekmann
Agata Ciabattoni	Beishui Liao	Marija Slavkovik
Robin Cooper	David Makinson	Jane Spurr
Luis Farinas del Cerro	Réka Markovich	Kaile Su
Esther David	George Metcalfe	Leon van der Torre
Didier Dubois	Claudia Nalon	Yde Venema
PM Dung	Valeria de Paiva	Rineke Verbrugge
David Fernandez Duque	Jeff Paris	Heinrich Wansing
Jan van Eijck	David Pearce	Jef Wijsen
Marcelo Falappa	Pavlos Peppas	John Woods
Amy Felty	Brigitte Pientka	Michael Wooldridge
Eduaro Fermé	Elaine Pimentel	Anna Zamansky

SCOPE AND SUBMISSIONS

This journal considers submission in all areas of pure and applied logic, including:

pure logical systems	dynamic logic
proof theory	quantum logic
constructive logic	algebraic logic
categorical logic	logic and cognition
modal and temporal logic	probabilistic logic
model theory	logic and networks
recursion theory	neuro-logical systems
type theory	complexity
nominal theory	argumentation theory
nonclassical logics	logic and computation
nonmonotonic logic	logic and language
numerical and uncertainty reasoning	logic engineering
logic and AI	knowledge-based systems
foundations of logic programming	automated reasoning
belief change/revision	knowledge representation
systems of knowledge and belief	logic in hardware and VLSI
logics and semantics of programming	natural language
specification and verification	concurrent computation
agent theory	planning
databases	

This journal will also consider papers on the application of logic in other subject areas: philosophy, cognitive science, physics etc. provided they have some formal content.

Submissions should be sent to Jane Spurr (jane@janespurr.net) as a pdf file, preferably compiled in \LaTeX using the IFCoLog class file.

CONTENTS

ARTICLES

- Normative Change: An AGM Approach 825**
Juliano Maranhão, Giovanni Casini, Gabriella Pigozzi and Leendert van der Torre
- Multi-agent Argumentation and Dialogue 891**
Ryuta Arisaka, Jérémie Dauphin, Ken Satoh and Leendert van der Torre
- The Law of Evidence and Labelled Deduction: Ten Years Later 925**
Dov Gabbay and John Woods
- Defeasible Deontic Logic: Arguing about Permission and Obligation 995**
Huimin Dong, Beishui Liao, Réka Markovich and Leendert van der Torre
- Business Process Modelling in Healthcare and Compliance Management:
A Logical Framework 1055**
*Ilaria Angela Amantea, Livio Robaldo, Emilio Sulis, Guido Governatori and
Guido Boella*

Explainable Reasoning with Legal Big Data: A Layered Framework 1079
*Grigoris Antoniou, Katie Atkinson, George Baryannis, Sotiris Batsakis,
Luigo Di Caro, Guido Governatori, Livio Robaldo, Giovanni Siragusa and
Ilias Tachmazidis*

Artificial Intelligence and Space Law 1105
*George Anthony Long, Cristiana Santos, Lucien Rapp, Réka Markovich and
Leendert van der Torre*

NORMATIVE CHANGE: AN AGM APPROACH

JULIANO S. A. MARANHÃO
University of São Paulo, Brazil
julianomaranhao@usp.br

GIOVANNI CASINI
CNR-ISTI, Italy, and University of Cape Town, South Africa
giovanni.casini@isti.cnr.it

LEENDERT VAN DER TORRE
University of Luxembourg
leon.vandertorre@uni.lu

GABRIELLA PIGOZZI
Université Paris-Dauphine, France
gabriella.pigozzi@dauphine.fr

Abstract

Studying normative change is of practical and theoretical interest. Changing legal rules pose interpretation problems in determining the content of legal rules. The question of interpretation is tightly linked to questions about determining the validity of rules and their ability to produce effects. Different formal models of normative change seem to be better suited to capturing these different dimensions: the dimension of validity appears to be better captured by the AGM approach, while syntactic methods are better suited to modelling how the effects of rules are blocked or enabled. Historically, the AGM approach to belief revision (on which we focus in this article) was the first formal model of normative change. We provide a survey of the AGM approach along with the main criticisms of it. We then turn to a formal analysis of normative change that combines AGM theory and input/output logic, thereby allowing a clear distinction between norms and obligations. Our approach addresses some of the difficulties of normative change, like combining constitutive and regulative rules (and the normative conflicts that may arise from such a combination), revision and contraction of normative systems, as well as contraction of normative systems that combine sets of constitutive and regulative rules. We end our paper by highlighting and discussing some challenges and open problems with the AGM approach regarding normative change.

1 Normative Change and Legal Reasoning

The study of normative change in identifying the law and understanding legal reasoning and legal interpretation is of practical and theoretical interest.

From a practical perspective, legal rules are the product of, or at least affected by, the continuous agency of authorities with the power to issue norms or make judicial decisions.¹ Such authoritative acts change the content of the normative order by including and excluding rules or by modifying their effects.

The problem lies in the fact that there are a variety of acts that perform such modifications in the lifetime of a normative system, which may have an effect on two dimensions:

- (i) **validity**: the pertinence of rules to a normative system that may be changed by acts of abrogation, explicit derogation or implicit derogation;
- (ii) **efficacy**: the capacity of rules to produce effects or apply in a certain time period, which may be changed by acts of annulment or invalidation, suspension, restriction, modulation etc.

Hence, there is a discrepancy between the period of the validity of a rule in a normative system (which also has its own time span of existence), and its period of efficacy, thus creating situations where a rule is invalid but applicable or where a rule is valid but inapplicable.

From a theoretical perspective, it is important to understand normative change in order to understand the status of entailed (derived) rules in a normative system and their relationship to explicitly promulgated rules. The debate about the status of entailed rules is connected to a central problem in the conception of modern law concerning the role of reason *versus* the role of authority in identifying the law [51]. The question is whether the ultimate basis for identifying the legal status of an action are considerations of moral correction or goodness, or determination by a social source, i.e. whether the legal status of an action is determined by the content of an authoritative act, which is objectively identifiable independently of moral or political arguments [62].

¹Even scholars like Dworkin [23] who refuse to reduce identifying the law to the content of authoritative social sources do acknowledge that those sources produce relevant legal material for legal interpretation, potentially affecting how the law is identified and causing modification to the law.

1.1 Normative Change and Legal Validity

The inclusion of a new rule in a normative system is performed by an act of promulgation (or enactment). This new rule may represent new content, changing the content of the normative system by making new obligations, permissions or prohibitions derivable. Or the new rule may be redundant, adding a new norm-formulation, new text, without actually introducing new content.

In turn, exclusion of a rule from the normative system or modification of its effects may be obtained by means of a variety of legislative or judicial acts. There are terminological variations and disputes concerning acts that either exclude content pertaining to normative systems or restrict its efficacy (applicability). There are also different practices depending on the jurisdiction, and particularly with respect to systems of common law vis-à-vis systems of statutory law. In order to avoid confusion, we shall use terms in accordance with their technical usage in legal practice, but will articulate their meanings where the terminology can be misleading. In general, we will use the terms *derogation* and *abrogation* to refer strictly to the *dimension of validity*, with the meaning that a statute is totally or partially excluded from (ceases to pertain to) the normative system. We prefer to restrict the term “annulment” to the dimension of efficacy, with the meaning that a rule or a set of rules has its effects cancelled (ceases to be applicable).

Derogation is a distinct normative act that excludes a rule or some rules from a set of valid rules. It may be explicit or implicit:

explicit derogation: a new rule that explicitly mentions the name of the rule or rules to be excluded.

implicit derogation: a new rule that adds normative content which is inconsistent with the content of previous rules in the normative system.

In the case of explicit derogation, the content of the new rule may consist of only excluding the named rule: for instance, “article 56 of Law 1234 is derogated”. In such a case, the derogation rule exhausts its effects by performing that very derogation [38].

Abrogation means excluding the totality of the rules of a statute. Usually, abrogation is due to an act of promulgating a new statute that substitutes the content of a previous statute on the same subject. The exclusion is explicit because the set of excluded rules is indicated by either naming the statute or indicating the subject-matter. Abrogation also introduces new content whose effects hold after the previous statute has been derogated.

Derogation and abrogation (as well as promulgation) are usually non-retroactive normative acts, producing their effects immediately after publication or at a certain time in the future indicated by the same act. In the legal jargon, their effects are *ex nunc*, i.e. “from now on”. That is, they are “established” by the legislative act.

We shall use the term “annulment”² to refer to acts that cancel the effects of a valid rule. If a rule is annulled, it becomes inapplicable, that is, one cannot derive obligations, permissions, powers or any legal consequences from it.

An annulment may be the consequence of a judicial declaration that a rule of the normative system is invalid, or it may be the product of legislative acts cancelling the effects of a rule. A judicial annulment recognises a “vice” or “defect” in the “pedigree” of the rule. Those “pedigree” defects are related to problems with the source of the rule, the *legitimate authority*, the procedure for creating the rule, or the incompatibility of the rule with the content of hierarchically superior rules. Depending on the gravity of the defect identified, the recognition may consider the rule to be invalid from the time of its promulgation (in the legal jargon, *ex tunc* effects) or from the moment the defect is declared (*ex nunc*).

To complicate matters, since the annulment may be a judicial act, the recognition of invalidity may be general, that is, applicable to all legal subjects, or it may have an effect on a particular legal relation or a particular individual. So there is a general dimension of effects, but there are also indirect effects where normative changes affect the legal positions of different individuals in different ways. The same also happens for derogation and abrogation, which cannot retroact, so that a derogated rule may still be applicable to facts that occurred before the derogation took place.

There are other ways to affect the efficacy of rules by authoritative acts. A statute or decree may suspend or restrict the applicability of a rule in a given period or to a given domain or context. For instance, the legal rules protecting moral rights for authors became inapplicable to software by the force of a new law (art. 2 §1 of the Brazilian Law 9609/1998 on Software Copyright). Or a rule may suspend the applicability of some rental of real estate or labour laws during a global pandemic.

Clearly, the temporal aspect is crucial to analysing normative change, and this temporal factor has two dimensions: the time span of the rule’s validity, that is,

²The term “revocation” is sometimes used in parallel with annulment and pertains to the dimension of validity. Revocation refers to the act of cancelling a previous declaration, contract or legislative act, but the term “annulment” is also used to refer to such a cancellation with the intent of producing legal effects, particularly when such a cancellation is performed by a different person or institution (e.g. a judicial court) to the one that issued the act (e.g. the parliament or the contracting parties). Annulment and invalidation may also refer to cancelling the *effects* or applicability of a particular act, and are therefore situated in the dimension of efficacy of rules.

the period of time in which the rule pertains to the normative system; and the time span of its applicability, that is, the period of time where the obligations/permissions derived by the rule are applicable.

These dynamics of normative change, which are performed by a variety of legal acts with different effects, bring a series of difficulties for determining the content and the effects of a normative system at a particular moment in time. Indeed, a promulgation and a derogation may involve choices between alternative and incompatible descriptions of the resulting normative system.

The practical import of the study of normative change is not only a matter of finding suitable formal and computable representations of an uncontroversial and standard practice. It is also relevant for clarifying that very practice by describing the impact of acts of promulgation and revocation on the content of a normative system, and especially how they affect the normative consequences or entailed rules of that system. We highlight three problems.

The first problem concerns the *network effects of normative change*, that is, the effects of a derogation or a promulgation on networks of regulative and constitutive rules [71]. Acts of promulgation or derogation may not only add or exclude *regulative rules*, which are authoritative rules demanding, prohibiting or permitting an action or the omission of an action. They may also add or exclude *constitutive rules*, whose role is to a) define under which factual conditions a certain object or action “counts as” an instance of a legal concept such as property right, or b) ascribe meaning to legal concepts via definitions (e.g. people under 18 years old count as minors).

Hence, stipulating a new definition or changing the definition of a legal concept may affect how the content of different regulatory rules are determined. In turn, the exclusion or addition of new rules that are related to a legal concept may affect the practical implications, and therefore the very understanding, of that very concept [68]. Such an effect is neither immediately nor completely acknowledged by lawgivers, and leads to subsequent modifications and adaptations.

For instance, the legal definition of “software” as “literary work”³ makes rules protecting the “expression” of a literary work applicable to the source code of software: the copyright owner may copy, share, or distribute the software, create “derivative work” etc. The equiparation also enhances new legal consequences by analogy, such as the additional copyright protection of the original “structure” of a code, considering that the “composition” of different non-original literary works are also protected. Thus, the addition of new rules or protections for “literary work” may also “expand” the protection of software. However, some undesirable legal consequences of that equiparation—for instance, the ascription of “moral rights” related

³Agreement on Trade-Related Aspects of Intellectual Property Rights (Trips Treaty, 1994)

to software, such as the right to regret and withdraw the work from distribution—have been derogated in several jurisdictions.⁴ Such derogations in turn affect the understanding of the very concept of copyright—originally conceived as intrinsically bound to the author’s personality—by linking the original notion of copyright to a network of personality rights. Thus, the ascription of new objects to a legal concept by definitional rules and the introduction or derogation of regulatory rules interferes with, and demands “reconfigurations” of, the links in the network of legal definitions and normative consequences.

The second problem concerns the *undecidability of implicit derogations*, which is a consequence of the potential conflict between different “collision criteria” in the law. New obligations, prohibitions, permissions or definitions added via lawgiving acts may create conflicts with the content of the previous version of the normative system. Such conflicts are solved by an *implicit derogation* operated by so-called *collision criteria*, which are legal principles of interpretation enunciating preference relations for solving conflicts between rules. There are three collision criteria:

lex superior: a *hierarchical* criterion according to which rules enacted by a source of a higher hierarchical degree prevail over rules from lower degree sources.

lex posterior: a *temporal* criterion according to which more recent rules take precedence over older ones.

lex specialis: a criterion of *specialisation* according to which a rule applicable to a specific circumstance or condition prevails over another rule applicable in a more general context.

Although it is clear that the hierarchical criterion prevails over the temporal and speciality criteria, the two last criteria may collide.

Example 1. *Suppose that a new statute on public concessions is promulgated stating:*

1. *A private company operating a public concession of a federal road may explore its margins for commercial purposes.*

This rule might conflict with a previous existing rule specific to electricity distribution companies stating:

⁴For instance, article 2º, §1, of the Brazilian Copyright Law considers all provisions of the law warranting moral rights to be inapplicable to software, except for the right to have authorship acknowledged and the right to oppose unauthorised modifications that may affect the reputation of the author.

2. *Public energy distribution companies have the right to use road margins to the extent that such use is necessary to install its energy transmission network.*

These rules conflict if one interprets the right to use, in which energy companies are invested, as the right to use free of charge, and if the right to explore the margins “for commercial purposes” is considered to include a right to charge a fee for the public energy distribution system. But the conflict cannot be solved by the existing collision criteria because there is a conflict in this case between *lex posterior*, which makes rule (1) prevail over rule (2), and *lex specialis*, which makes rule (2) prevail over rule (1). Actually, there is another possible source of dispute, which is the understanding of which rule is the more specific rule. One could argue that rule (2) is more specific because it relates to a public energy distribution company, while rule (1) relates to all kinds of potential users. However, one could also argue that rule (1) is more specific because it relates to roads, the object of public concessions to private companies, while rule (2) has a wider scope on this aspect.

Hence, given a conflict of rules created by a promulgation, there may be no fixed criteria for deciding which one should prevail.

The third problem concerns the *indeterminacy of implicit derogations*, that is, that the promulgation of a new rule may conflict with a rule derived from the combination of different explicit rules in the normative system.

Example 2. *Suppose that a regulation contains the following rules:*

3. *Brasilia is the capital city of the Brazilian Federation.*
4. *The Brazilian Federal Administration must be located in the capital city of the Brazilian Federation.*

Now suppose that the following rule is promulgated:

5. *The Brazilian Federal Administration must be located in Rio de Janeiro.*

Rule 3 does not conflict with either rule 1 or 2, but it does conflict with the entailed rule:

- 5'. *The Brazilian Federal Administration must be located in Brasilia.*

This would be a case of *implicit derogation* of an entailed rule resolved by the temporal criteria of collision. However, the entailed rule can only be suppressed if at least one of explicit rules (3) or (4) are derogated. Hence, the content of the

normative system after the promulgation of (5) is undetermined, with three possible candidates for the outcome of this derogation: $S_1 = \{3, 5\}$, $S_2 = \{4, 5\}$ and $S_3 = \{5\}$.

From a domain-specific consideration, S_1 is plausible although it may have perplexing consequences (for instance, if there is a rule assigning a budget to the Brazilian capital that includes expenses for relocating and maintaining the offices of the Federal Administration). System S_2 would not properly imply that:

- 3'. Rio de Janeiro is the capital city of the Brazilian Federation.

But promulgating a norm specifying a city other than Rio de Janeiro as the capital city of Brazil would again lead to inconsistency.

Finally, system S_3 would leave the capital city of Brazil undefined, which could create uncertainty in the application of other rules employing that concept.

A similar problem of indeterminacy would appear when a rule entailed from a new and hierarchical superior rule is promulgated.

Example 3. *Suppose that a normative system contains the following rule:*

6. *All industries are free economic activities except for the public services listed below: (...)*

Suppose that the aviation industry is not listed in rule (6), implying that aviation is a free economic activity, and suppose also that there is a federal statute (the Aviation Code) stating the following:

7. *Aviation companies must be controlled by national investors.*

Now consider that a constitutional rule is enacted imposing the following:

8. *There ought to be no discrimination between the national and foreign capital of companies dedicated to any free economic activity.*

Considering that control by national investors counts as “discrimination” between foreign and national investors, rule (8) conflicts with rules (6) and (7), although originally the last two rules seemed to have no relevant connection to each other. The inconsistency is solved if either of these last two rules is derogated. The first option is to delete constitutive rule (6), which classifies the aviation industry as a free economic activity. The second option is to delete rule (7), thereby weakly permitting, that is not prohibiting, the control of aviation companies by foreign investors.

Hence, the interaction between constitutive and regulative rules, the problem of implicit derogation and the derogation of entailed rules all open up different possibilities for identifying the normative system resulting from normative revisions. Logical analysis of normative change should be faithful to such an indeterminacy, making the different possibilities for the resulting normative system transparent. Legal interpretation and argumentation may provide further constraints in order to select which, among all the possible candidates, would be the preferred outcome of a derogation, which may be domain-specific, or may have its rationality represented in formal models of normative change.

1.2 Normative Change and Legal Interpretation

Legal reasoning can be conceptually structured as three main tasks, as suggested by Wroblewski [77, 78]:

- (i) **validity**: identifying the valid legal rules that are generally applicable to the subject-matter;
- (ii) **interpretation**: determining the content of the rules identified as valid;
- (iii) **application**: instantiating the content of the valid rules applied to concrete or hypothetical cases (this last task includes identifying the relevant facts of the case, identifying how they qualify according to the applicable rules, and determining the legal consequences based on those rules).

At first glance, normative change should only be concerned with questions of validity, since the dynamics of promulgation and derogation determines the timeframe for the applicability of rules in normative systems. However, the three problems highlighted above show an intrinsic connection between normative change and legal interpretation, given that one of the main triggers of normative dynamics is the need to handle inconsistencies between the *content* of different rules in the normative system.

The problem of *network effects* is connected to determining the content of regulative rules with conceptual definitions. The *undecidability* problem is also about choosing between rules with conflicting content. The *indeterminacy* problem of implicit derogation concerns a conflict between the content of the promulgated rule and the content entailed by the normative system.

Given that the core task of legal interpretation is to determine the content of legal rules, it is necessary to first identify inconsistencies between rules, and therefore to check whether an implicit derogation has undermined the validity of a rule. Hence,

questions of validity and interpretation are not serial but circular. The object of interpretation is the content of valid rules, but interpretation is also necessary to the inquiry about validity. The same applies to interpretation and application. Since the conditions for applying the rule may not be isomorphic to the factors or circumstances of the case at hand [60, p. 77 ff.], the rule must be adapted to become “operational”. Further qualifications to the facts must be introduced via definitions that match the factual properties of the case with the concepts employed in the rule in order to make them isomorphic [1]. Hence, although it is the content of the rule that is subsequently instantiated, that instantiation induces modifications to the content of the rule to be applied [66, p. 36 ff.].

Hence, interpretation is pervasive in legal reasoning, performing an important role from identifying the authoritative sources to determining the legal effects on a concrete or hypothetical case.

Broadly understood, legal interpretation encompasses both *linguistic* and *constructive* interpretation. Linguistic interpretation consists in identifying the semantic/pragmatic content that is conveyed by an authoritative legal text.⁵ In turn, constructive interpretation, or “legal construction” [72], consists in determining the legal effect of that linguistic content, which means constructing the content of an “operational rule”.

Some conceive of linguistic interpretation as an inquiry into the linguistic facts of a language community [11, 72, 54], while others include an evaluative component in every linguistic inquiry [26, 23], and therefore consider the whole process of interpretation as constructing rules in the light of the purpose of legal practice. But even those who question the distinction accept that there would be a pre-interpretive stage where some preliminary meaning ascription takes place.

The linguistic interpretation or pre-interpretive stage may provide unsatisfactory solutions for a particular case. The linguistic meaning of the rule may not indicate a normative solution to a particular constellation of relevant facts [4], leaving a so-called “gap” in the normative system that must be fulfilled. The linguistic inquiry may also provide conflicting commands deriving from the same rule or from different rules, in which case the contradiction must be corrected. It may provide an array of alternative meanings (ambiguity), from which only one must be chosen, or may provide an imprecise meaning (vagueness), demanding further definitions to determine whether the case at hand fits the conditions for applying the rule. Finally, the rule’s command as determined by the linguistic inquiry may violate the rule’s underlying justification (the values promoted by the rule), which may necessitate

⁵Legal theorists disagree about what is the object of legal interpretation. While some contend that the object of interpretation is to formulate norms from authoritative sources [64], others, like Dworkin [23] would also include the whole argumentative social practice of law [22].

the introduction of exceptions or the specification of new conditions for applying the rule so that its content aligns with its purpose.

These further processes of

- filling gaps by adding new content,
- eliminating ambiguities by choosing between different content,
- eliminating vagueness by adding definitions to make the rule precise,
- resolving inconsistencies between rules by excluding content, and
- resolving deviances to the rule's command with respect to its underlying justification by modifying its conditions of application,

all clearly involve changes not only to the rule to be applied but also to the very normative system. The process of constructing an operational rule to be applied presupposes that the interpreted rule coheres with the normative system, and therefore that what is instantiated is actually a reconstructed fragment of a normative order containing a set of rules that are relevant to defining the deontic status (obligatory, forbidden, permitted) of the action at stake [4]. This reconstruction may be performed by a judge to solve a concrete case (judicial interpretation), or in legal doctrine when indicating solutions to hypothetical legal cases (doctrinal interpretation).

Note that in practice it is difficult to discriminate between these two different dimensions of legal interpretation—linguistic and constructive—considering that the very ascription of meaning to legal texts is constrained by a presumption of the lawgiver's rationality or "unity of will" [14], which requires that a text must be given a meaning that avoids inconsistencies or misalignments with the rule's purpose, and preferably avoids gaps and imprecision. Hence, construction may take place even when the identification of the meaning of a rule is uncontroversial.

For instance, consider the regulation on abortion in the Brazilian Criminal Code.

9. Causing an abortion; Punishment: imprisonment from 1 to 3 years.

10. Abortion performed by a physician is not punishable: (i) if there is no other way to save the pregnant woman's life; (ii) the pregnant woman has consented to the abortion and the pregnancy is the result of sexual abuse.

A criminal lawyer would say that it is settled from the text above that it is forbidden to abort if the pregnant woman's life is not endangered and no sexual abuse took place. Some would even say that this conclusion is immediate and does not

require interpretation. However, first of all, the interpretation of clauses (i) and (ii) as disjunctive and not conjunctive involves some evaluative considerations favouring women's freedom. Secondly, the plain language meaning actually reveals inconsistency between rules (9) and (10). Rule (10) is read as an exception, but this means that some interpretation cannons operate in order to first assume that inconsistent rules should be applicable to different hypothetical conditions, then to derogate (9) by specificity, and finally to reintroduce the prohibition of causing an abortion in scenarios that have not been exempted (*exceptio firmat regulam in casibus non exceptis*). The "operational rules" reconstructed from the original linguistic meaning are thus:

- 9*. Abortion is forbidden if not performed by a physician or if there are other ways to save the pregnant woman's life and the pregnancy is the result of sexual abuse or the pregnant woman has not consented to the abortion.
- 10*. Abortion is permitted if performed by a physician and there is no other way to save the pregnant woman's life or if the pregnancy is the result of sexual abuse and the pregnant woman has consented to the abortion.

The fact that what is assumed to be the "plain language meaning" of a norm already involves its construction leads some to consider the object of legal interpretation to be the legal community's set of settled instantiations of the valid rules [54] rather than the ordinary meaning of legal texts. In this conception, legal interpretation would then be the process of construction from that restricted basis of settled law, in order to develop solutions for unclear cases with gaps, imprecision and/or conflicts, etc.

Legal construction allows flexibility in the law so that it can adapt to new circumstances and social demands while reinforcing the authority of the normative order. It can achieve this by keeping track of the original rules (taking as a starting point the legal text, the clear and settled instantiations, or the legal history) and making them align with community values. Assessment of this interpretative practice from the perspective of normative change reveals different strategies used in legal doctrine, or by the courts, to manipulate the legal material in the sources in order to justify choosing a particular legal solution. Particularly interesting is their stipulation of definitions affecting relevant concepts of the rule.

Consider, for instance, the controversy in many jurisdictions concerning police access to the content of mobile phones in search & seizure orders.

In 2014, a decision by the Brazilian Superior Court of Justice (STJ: HC 51.531-RO) held that a WhatsApp conversation on a mobile phone collected in a search

procedure is analogous to ongoing correspondence and should count as “written communication”. Therefore, an *order to intercept* was mandatory to access its content, otherwise the access would have violated freedom of communication. However, in a decision reached in 2016 (STJ: HC 75.800-PR), the same court affirmed that a message exchange on a mobile phone is just stored data and therefore a property item which, according to the statutes, may be accessed in a search & seizure procedure.

The German Constitutional Court (BVerfGE, 115,166, *Kommunikationsverbindungsdaten*) also concluded that access to data stored on a mobile phone collected during an investigation does not violate rules regarding search & seizure. Such data would be analogous to information in a physical document since both involve possession and the data or information could have been destroyed by the searched individual. Therefore, accessing the history of calls does not affect freedom of communication, and does not have a greater impact on informational autonomy or property rights deserving special protection.

Example 4. *Consider a normative system with the following regulative rules:*

11. *Police officers have the power to access any property item if and only if authorised by a judicial search & seizure order.*
12. *Police officers have the power to intercept written or oral communication if and only if authorised by a judicial interception order.*

The following conceptual rules are key to determining whether stored text messages may be accessed in a search & seizure order:

13. *A message exchange stored on a mobile phone counts as ongoing communication;*
14. *A message exchange stored on a mobile phone counts as stored data;*
15. *Stored data counts as a property item.*

Suppose that officers only hold a search & seizure order. Then there is an inconsistency between conceptual rule (13), on the one hand, and conceptual rules (14) and (15) on the other. The difficulty lies in the fact that the linguistic meaning of a message exchange supports its qualification as both communication and stored data. The link between stored data and property pertains to the legal language and derives from valid legal rules. The German court has just excluded rule (13), thus avoiding that the search procedure should become unconstitutional by affecting freedom of communication. One of the Brazilian courts chose to delete rule (14).

But those qualifications (data as property, stored messages as data, message exchanges as communication) are also relevant to the application of other rules. Another solution to keep rule (11) compatible with the constitutional value of freedom of communication, and with a lower impact on the network of conceptual and regulative rules, would be to refine rule (11) as follows:

- 11*. Police officers have the power to access any property item, except for the digital content of mobile phones, if and only if authorised by a judicial search & seizure order.

Indeed, this was the solution adopted by the U.S. Supreme Court in a similar case involving search powers in an arrest (*Riley v. California*, 2014).

Hence, legal construction involves manipulating conceptual definitions not only by legal doctrine, but also regulative rules. This possibility does not offend the authority of the rules provided that, first, conceptual definitions may also be stipulated by valid legal rules, and secondly, that valid regulative rules may be derogated or refined by introducing exceptions, in the name of consistency with constitutional values, as explicit and higher order rules [8].

But it is clear that legal construction and legal interpretation in general have both a conservative and a creative component [22]. On the one hand, construction must be faithful to the settled normative order. On the other hand, it must enhance new solutions by clarifying the content of that order. In other words, choices and changes to the content of the legal order are going to take place, but only to the extent that is minimal and necessary to clarify its content.

It is also characteristic of such constructions that their conclusion is presented as entailing a coherent interpretation of the normative system. Opposing conclusions in apparently similar cases are shown to align with the balance of the relevant values pursued by the normative system. Alignment is attained by using an array of different techniques in constructive interpretation: discarding possible conceptual qualifications e.g. excluding the rule that stored messages count as communication, introducing exceptions to rules e.g. excluding mobile phones from the general search powers of officials, and introducing or excluding values from consideration.

It is clear from this discussion and examples that legal construction as a fundamental dimension of legal interpretation consists in making changes to the content of the normative system, and that these changes are driven by both a demand for coherence and by a demand for conservatism or “minimal change” to the legal order. These drivers show how logics of theory change are suitable for modelling legal construction.

To conclude this practical perspective, we observe that the relationship between interpretation and normative change is twofold. On the one hand, legal interpreta-

tion is a precondition to the dynamics of normative systems, as the identification of inconsistencies between the content of rules depends on it. On the other hand, the very activity of legal interpretation may be seen as dynamics of change affecting constitutive and regulatory rules.

1.3 Normative Change and Implied Rules

From a theoretical perspective, normative change is an important factor in understanding the status of implied (derived) rules in a normative system and its relation to explicitly promulgated rules. The debate about the status of entailed rules is connected to a central problem in the conception of modern law concerning the role of reason *versus* the role of authority in identifying the law. The question is whether the ground for identifying the legal status of an action consists in reasoning about its correction or goodness or whether this status is determined by the will of an authority with respect to individual or collective behaviour or its outcome.

If one conceives that the binding force of the content of explicit rules is the outcome of the authority's will manifested in the norm-giving act, the question arises whether or to what extent obligations, prohibitions or permissions deductively derived from those original rules, albeit not explicitly endorsed by the authority, are also binding or should also be considered to be part of the normative system.

This problem may be explored from the perspective of normative dynamics. Instead of a synchronic epistemology considering the identification of a rule as a matter of examining the foundational or coherentist connection of its content to the content of the other rules of the system [10], one may adopt a diachronic perspective of examining the vulnerability of the rule's content to changes in the normative system. If derived rules have the same "ontological status" as explicit rules, then, on the one hand, the promulgation (addition) of derived rules would be redundant and, on the other hand, their derogation would immediately mean a change in the normative system.

For instance, the Brazilian Criminal Code forbade sexual abuse with the following set of explicit rules:

16. It is forbidden to practice sexual intercourse without consent.

17. Sexual intercourse with a person under 14 years old shall be considered to be without consent.

Should we consider the derived rule (18) below a valid legal rule of the Brazilian criminal law system?

18. It is forbidden to practice sexual intercourse with a person under 14 years old.

A decade ago, a controversial decision by the Brazilian Supreme Court ruled that habeas corpus applied to an offender who maintained a sexual relationship with a 12 year old girl. The legal community has interpreted that ruling as *contra legem*, since it was widely assumed that the act violated the criminal code. It seems plain enough that although rule (18) was not explicitly promulgated, compliance with its content should be obligatory and any disregard would be a violation. And this follows from the fact that the content of (18) is deductively derived from rules (16) and (17).

Given that there is such a derived obligation, some would argue that rule (18) is also part of the normative system [4, 55]. Here, the binding force of the obligation is an outcome of reasoning (deduction), and if law is the system of binding rules, it should be part of the normative system as well.

Some, however, would accept the binding force of such derived rules, but would not acknowledge them as part of the normative system if their content is not explicitly willed [54]. Accepting them as part of the normative system, Marmor argues, would imply a (most probably) false assumption that the set of legal rules is coherent. Others, like Joseph Raz [63], would only accept them if such derivations were endorsed by the relevant authority (even though it is not quite clear what such endorsement means) as something distinct from explicitly willing its content but inferring such content from the explicit rules.

Curiously enough, that controversial decision by the Brazilian Supreme Court led to a legislative act (Law 12.015/2009) introducing rule (18) as an explicit rule of the Code. Did that law effectively change the Brazilian criminal law system? One could say that these are two different formulations of the Code representing the same criminal law system, provided that they contain the same set of derived obligations. If this is true, what led to the promulgation of the new legislative act?

One could say that it was fundamentally a political gesture with redundant or irrelevant legal consequences. Or one could say that the Supreme Court had actually changed the law, which was later modified by legislation again. But the interesting question is: if two different normative systems have identical normative consequences, is it the case that identical promulgations or derogations in each of these systems would lead to the same resulting normative system?

Example 5. *Consider normative system S_1 with the following formulations:*

16. *It is forbidden to practice sexual intercourse without consent.*

19. *Sexual intercourse with a legally incompetent person shall be considered to be without consent.*

20. *A person becomes legally competent by reaching 14 years of age.*

Now consider normative system $S2$ containing rules (16), (19), (20) and, in addition, (18) as an explicit rule.

18. *It is forbidden to practice sexual intercourse with a person under 14 years old.*

Suppose now that the following rule is promulgated:

21. *A person becomes legally competent by reaching 16 years of age.*

Clearly, rule (18) is derived from $S1$. Hence, from the synchronic perspective, it is clear that $S1 = S2$, since the set of derived obligations is the same. But the effect of promulgating rule (21) in $S1$ is different from its promulgation in $S2$. In $S1$, promulgated rule (21) substitutes rule (20), and therefore the revised system ($S1^*$) derives the following:

(18*) It is forbidden to practice sexual intercourse with a person under 16 years old.

However, in system $S2$, rule (18) would still be derived. And while rules (20) and (21) conflict, this is not necessarily a conflict between explicit rule (18) and derived rule (18*). Therefore rule (18) could still be derivable. It would be a matter of legal interpretation to determine whether the new definition of legal competence would be applicable only to civil law, that is, the ability to perform valid civil and contractual acts, or whether it would also be applicable to criminal law, specifically, the ability to consent to sexual intercourse or to be liable to criminal responsibility.

Hence, from a synchronic perspective, i.e. considering the normative system at a particular moment in time, one may assume that two normative systems are the same if they derive the same set of obligations/permissions, even if they have different formulations. That is, from that perspective, the formulation of the base of explicit rules is irrelevant. However, from a diachronic perspective, that is, considering the normative system's change from one moment to a second moment where a new rule is promulgated or derogated, the formulation of the base of explicit rules becomes relevant, given that the revision of different sets of explicit rules with the same derived obligations/permissions may lead to different outcomes. Therefore, changes in the base of explicit rules may not result in changes in the set of obligations/permissions, but every change in the set of obligations/permissions means a change in the base.

This observation makes it clear that even if one assumes that the content of derived rules is as equally binding as the content of explicit rules, which would make these rules share the same “normative status”, it is not the case that they should share the same “pertinence status”. That is, the fact that a derived obligation is binding does not imply that it is a rule pertaining to the normative system.

1.4 Modelling Normative Change

The distinction between the dimension of the validity of a rule (the time span of the pertinence of a rule to the normative system) and the binding force or efficacy of derived obligations or permissions (the time span where obligations and permissions are applicable) is also relevant for defining an appropriate methodology for the study of normative change. The different methods may focus on one or another aspect of normative change, namely, changes to the content of norms that are part of the normative order, or changes with respect to the effectiveness of obligations over time.

Suppose that there is a normative system S_3 with the rule:

22. Abortion is forbidden.

Since this is an absolute prohibition, it applies to every possible circumstance. Therefore, the following prohibition is derived:

23. Abortion is forbidden if the pregnancy is the result of sexual abuse.

Suppose that a legislative or judicial authority wants to change rule (23) by permitting abortion in the case of sexual abuse (or a legal scholar argues that there is an “implicit exception” to the prohibition of abortion based on the constitutional value of a woman’s dignity). This normative change may be described in at least three different ways corresponding to three different methods proposed in the literature on artificial intelligence & law for modelling normative change.

The first methodology, devised by Governatori and Rotolo [30], may be called the *syntactic approach*. According to this approach, norm change is an operation performed on the rules contained in the code for determining whether a default rule is applicable or ceases to be applicable in defeasible deontic logic. So, the focus of the approach is not really the dimension of validity (the pertinence of the rule to the normative system) but the dimension of the efficacy (applicability) of derived obligations and permissions. They call “annulment” the operation where all the past and future effects of the rule are cancelled and “abrogation” the operation where

only the effects to the future are cancelled while past effects still hold. They use a temporal extension of defeasible logic to keep track of changes in the normative system and to deal with retroactivity (the possibility of changing the applicability of obligations and permissions in the past). As we have seen, there are two temporal dimensions to be tackled: the time a norm is valid (when the norm enters the normative system) and the time it is effective (when the norm can produce legal effects). As a consequence, multiple versions of the normative system are needed [30].

The logical machinery used to represent normative change in this approach is complex given that the default logic has to gather very different sorts of default rules providing information on: the content of rules, meta-rules regarding the applicability of other rules, preference between rules, and the timeframe of applicability. For instance, an “abrogation” of a default rule is represented by the addition of a defeater, which is a default rule of a higher order with void content, that is, from which no obligation or permission is derived.

For the example on the regulation of abortion above, the syntactic approach could be roughly illustrated by indicating that in the case of sexual abuse, rule (22) is *not applicable*, and therefore rule (23) is not derived. This could be achieved by introducing a sort of meta-rule to the normative set stating:

- 24. In the case of sexual abuse, rule (22) is not applicable.

Such a rule would be a *defeater* because it would block the derivation of consequences from rule (22) without excluding it from the normative system. Notice that it adds no normative content by itself.

It is also possible to strengthen the contention that abortion is permitted in the case of sexual abuse by adding another rule to the normative system stating:

- 25. Abortion is permitted if the pregnancy is the result of sexual abuse.

In Governatori and Rotolo’s approach, this addition is obtained by turning a defeater into a default rule that blocks the application of the original prohibition, but also derives the content of a permission in the case of sexual abuse.

This representation, however, does not capture the basic intuition that derogation is a sort of exclusion where the rule ceases to be a part of the normative system. Instead, since the model concerns the dimension of the efficacy of obligations, a derogation is captured only by blocking the effects of a default rule. Besides, what can be derived depends on which rules are valid at the time when we do the derivation. Thus, in order to keep track of norm changes, Governatori and Rotolo represent different versions of a legal system.

In order to reduce such complexity, Governatori *et al.* [31] explored three AGM-like [6, 7] contraction operators to remove rules, add exceptions and revise rule priorities. Governatori *et al.* [29] also explored a model where, on particular occasions, normative change is reduced to a change of preference relations between default rules.

To illustrate this second method, which may be called the *preferential approach*, consider that from a moral order or a set of constitutional values one may derive inconsistent standards regarding abortion. One may derive permission of abortion from moral considerations, or from arguments about constitutional values, regarding the axiological contention that “women are free to dispose of their own bodies”. But one may also derive prohibition of abortion (rule 22) from a moral contention, or from a constitutional value, stating that “all human beings are the subject of moral worth” and the determination that a “foetus is a human being”.

Hence, this normative system would include rule (22) as well as the following rule:

26. Abortion is permitted.

The presence of rules (22) and (26) makes the normative system inconsistent, and thus the determination of the consequences of these conflicting rules for each relevant circumstance would depend on the addition and change of preference rules such as:

27. In the case of sexual abuse, rule (26) is preferred over rule (22).

In these two alternatives for representing change (syntactic and preferential), the corresponding logic cannot be classical (in particular, it cannot be monotonic). Otherwise rule (22) would conflict with rule (25) and rule (26), thereby making the normative system trivial. In these descriptions, rules (22), (25) and (26) are part of the normative system as “defaults”, and there may be circumstances where each of these becomes inapplicable, or where one of them prevails over another. With the syntactic approach, normative change is a matter of adding new defaults or defeaters to block or enable the normative effects of the defaults over time and according to relevant factors or circumstances. With the preferential approach, normative change is reduced to changing the preference relations between default rules on particular occasions.

In both the syntactic and preferential approaches, a change in the normative system should include not only information about the content of the rules that are subject to change but also information about the applicability of these rules. It is this information about applicability and preference that determines the set of

obligations and permissions derivable from the normative system. Actually, in both these approaches, the set of obligations and permissions may change without any modification to the content of the rules belonging to the normative system. It may be the result of modification to the time span of the applicability of the rules in that set, or the result of a change in the preference relations between defaults.

A third approach, which may be called the *AGM approach*, represents derogation and enactment, respectively, as effective exclusions and additions of content to the normative system. Historically, this was the first approach to modelling normative change, and was originally proposed by Alchourrón and Makinson [6, 7]. When Gärdenfors joined (at that time he was mainly working on counterfactuals), the trio became the founders of the well-known AGM theory, and started the fruitful research area of belief revision [5], which has found many applications in computer science and epistemology. Belief revision is the formal study of how a theory (a deductively closed set of propositional formulas) may change in view of new information that may cause inconsistency with existing beliefs. The basic operations of belief change are expansion (which corresponds to the promulgation of a rule to a code), revision (which corresponds to amendment of the code) and contraction (which corresponds to derogation of its normative application).

One of the first attempts to specify the AGM framework to tackling normative change was put forward by Maranhão [46, 47]. Maranhão introduced a *refinement* operator, which restricts the acceptance of new input to certain conditions in a revision, or keeps a more refined (weaker) version of a rule to be excluded in a contraction. Refinement thus represents the introduction of exceptions to rules in order to avoid conflicts in normative systems (see section 3.6).

More recently, Boella *et al.* [16] also reconsidered the original inspiration for the AGM theory of belief revision as a framework for evaluating the dynamics of rule-based systems. They observed that if we wish to weaken a rule-based system from which we derive too much, we can use the theory of belief base dynamics [34] to select a subset of the rules as a contraction of the rule-based system. Base contraction seems to be the most straightforward and safe way to perform a contraction; it always results in a subset of the original base. But it sometimes means removing too much. In turn, AGM theory contraction may retain some implications of the rule to be deleted. This was one of the motivations for the present contribution. Another advancement is to represent normative change in a formal framework that clearly distinguishes between the concepts of the pertinence of a rule in a normative system and the effectiveness of an obligation in a given context using the input/output logic framework developed by Makinson and van der Torre [42]. A similar approach was proposed by Stolpe [73]. In that work, AGM contractions and revision are used to define derogation and amendment of norms. In particular, the derogation operation

is an AGM partial meet contraction obtained by defining a selection function for a set of norms in input/output logic. Norm revision defined via the Levi Identity characterises the amendment of norms. Stolpe can thus show that derogation and amendment operators are in one-to-one correspondence with the Harper and Levi Identities as inverse bijective maps (cf. section 2.1). Also, Tamargo *et al.* [74, 75] recently studied AGM-like revision operators that consider rules indexed by time intervals.

In the AGM approach, the operation of normative change is performed on the normative system (the set of rules that may be closed under logical consequence). The rules in the original system or in the system resulting from change does not carry meta-information about their applicability, time span or hierarchy (although these features may be added). Therefore, the set of applicable obligations or permissions at a given moment in time is the set of all logical consequences of the normative system valid at that specific time. Hence, information about hierarchy and the time span of validity and applicability is not part of the representation of its rules and does not interfere with the derivation rules of the underlying logic (although such information might be relevant to the revision functions).

To illustrate the AGM approach to the example of abortion discussed above, the normative change would consist in refining rule (22) with respect to the defeating factor “pregnancy resulting from sexual abuse”, resulting in a normative system where rules (23) and consequently (22) are deleted and containing the following rules:

25. Abortion is permitted if the pregnancy is the result of sexual abuse.
28. Abortion is forbidden if it is not the case that the pregnancy is the result of sexual abuse.

With this last approach, every normative change, that is, every change in the set of obligations and permissions derived from the normative system, amounts to a change to the content of the rules that belong to the set of norms. This aspect makes the set of obligations and conditions for their application closer to the content of the revised normative system.

Research on formal models of normative change has also been concerned with representing legal interpretation.

In the field of artificial intelligence & law, legal interpretation has been mainly explored with models of case-based reasoning, where teleological reasoning is represented to derive solutions to new cases based on precedents. Following Berman and Hafner [13], AI & Law research on teleological reasoning has provided multiple models of the relationship between cases, the factors that such cases include or express,

and the values at stake. Bench-Capon and Sartor [12] assign values to factors, and consequently to rules embedding such factors, to explain precedents according to the applicable rules and the importance of the values promoted by such rules. Prakken *et al.* [61] formalise teleological reasoning using logics for defeasible argumentation, extended to allow the possibility of expressing arguments about values, supported by cases. Sartor [69] explores the proportional balance of constitutional rights, where a legal outcome is compared to alternative outcomes based on their impact on the promotion and demotion of values. He examines the level of consistency between value-based decisions of cases given the factors present in those cases [70].

In turn, AI & Law research on statutory interpretation has focused on the dynamic ascription of meanings to rules. These contributions are based on the distinction between “*constitutive*” (or “*conceptual*”) rules ascribing meanings to facts or objects and “*regulative*” rules demanding, prohibiting or permitting actions or states [32]. Interpretation is then modelled as introducing or changing conceptual rules. Governatori and Rotolo [30] represent such changes, within the syntactic approach, as the introduction of exceptions, by blocking the application of default rules to a given condition or constellation of factors. Boella *et al.* [15] developed that model by introducing values as coherence parameters guiding the change of conceptual rules, parameters whose meanings may be extended (weakening the antecedent of a conditional rule) or restricted (strengthening the antecedent of a conditional rule).

The incorporation of the AGM approach into input/output logics [16] and, later, the representation of normative systems in an architecture of input/output logics combining constitutive and regulative rules, brought a new perspective to representing legal interpretation [18]. Maranhão and de Souza [52] introduced a contraction function for such combined normative sets in order to represent choices in legal doctrine between changing the definitions (or meaning ascriptions) of legal terms and changing the content of legal regulative rules, taking into consideration the network effects of those changes.

Maranhão [50] proposed an architecture of input/output logics for modelling doctrinal interpretation where values are represented as rules, and constitutive and regulative rules are the object of different contraction, revision and refinement functions. Differently from the work of Boella *et al.* [15], where legal interpretation is conceived as a dynamic of syntactic modifications to constitutive rules (within the syntactic approach), in Maranhão’s model it is not only constitutive rules, but also values and regulative rules, that are subject to change (with the AGM approach) in order to reach a coherent and stable description of the normative system. More recently, Maranhão and Sartor’s [53] research on statutory interpretation built on the case-based tradition of teleological reasoning and balancing with their repre-

sentation of legal construction—where a model of balancing values is incorporated into an architecture of input/output logics—serving as a reference to the revision of constitutive (meaning ascriptions) and regulative rules.

Which is the best approach to representing normative change—syntactic, preferential or AGM?

This question was controversial in the 1990s in the context of Alchourrón's [3] criticism that defeasible logics are philosophically inadequate. According to Alchourrón, defeasible logic unnecessarily weakens the inferential power of the underlying logic. It obscures the fact that the defeat of a conclusion is actually the result of the dynamic of revising the premises in a derivation, or the fact that the defeat of a consequence results from revising the antecedent of a conditional. According to Alchourrón, in an adequate account of the epistemology of law or of any domain, the revision processes of the premises of an argument or the antecedent of a conditional should be transparent [48].

Actually the reply to this question depends on what aspect of legal reasoning one would like to capture with the model of representation (without considering the technical issue of computational complexity).

As we have seen, there is a fundamental difference between the pertinence of a rule to a normative system and its effects in terms of the derivability of the corresponding obligations/permissions in the presence of given circumstances. There is the time span for when a rule pertains to the normative system, that is, the time the rule exists in the normative system. But, although pertinent to a system, a rule may still not produce its effects, for example because its conditions of application are dependent on an event or regulated by another rule, so there is another time span for when the norm is applicable. Furthermore, as mentioned above, there is the time span for when the conclusions of an instantiated rule apply to a particular individual, considering that the instantiated rule may be derogated or annulled (i.e. declared invalid) for that particular individual by a judicial authority.

The distinction between the validity and efficacy of a rule may be captured by all approaches. But the syntactic approach seems to be more congenial to the dimension of efficacy, that is, the applicability of rules, considering that the revision operations are represented as syntactical changes to the rules that affect their applicability. A contraction operator does not properly exclude a rule but interferes with the derivability of its consequence.

In turn, the AGM approach seems to be more congenial to modelling the dynamics of the pertinence of a rule in a normative system, since the suppression or addition of obligations or permissions, and obligations derived from the basic set of rules, are reflected in proper exclusion or expansion to the rules of the normative system.

In the end, the description of the obligations and permissions derived from the normative system may coincide in both approaches, the difference lying in the set of basic rules.

Lastly, the preferential approach seems to be more congenial to the dynamic of legal principles and values related to positively enacted rules. Such principles and values, both considered as external to the normative system or enshrined in the constitution, potentially conflict but coexist in the normative order or political morality underlying such an order of legal rules. Depending on the context, they are balanced in order to derive a solution. The preferential approach reflects the fact that the derivation of a normative solution from principles or values results from resolving potential conflicts by giving more weight to a preferred principle than another principle in a given context.

It seems that a closer correspondence between the content of the rules and the applicable obligations/permissions is also of interest for the representation of legal construction where a particular reconstruction of a fragment of the normative system takes place before the instantiation of an operational rule.

Recent research on models of legal interpretation has shown that the three approaches must be combined since, as we have seen, the interpretive activity, particularly legal construction, involves all of the following three dimensions:

- manipulation and refinement of constitutive and regulative rules in a normative system (*validity*);
- consideration and weighing of underlying values (*balancing*);
- adaptation of definitions of legal terms to make the rules isomorphic and applicable to the facts of a particular case (*applicability*).

The first two approaches listed in this section are presented in the work of Tamargo *et al.* [75]. This article focuses on the AGM option, presenting its reformulation for input/output logics—a family of logics dedicated to the analysis of normative reasoning in particular as well as rule-based reasoning in general. We consider the combination of these two formal approaches, AGM belief change and input/output logics, to be a promising framework for analysing normative change. On the one hand, the kind of analysis of information change that AGM-like approaches pursue is insightful and very clear at the same time, and often can be reformulated into specific solutions for other formal frameworks. On the other hand, input/output logics offer an analysis of rule-based reasoning that is along the same lines, since it combines the immediate clarity of characterising distinct rule-based systems via the structural properties they satisfy with an in-depth analysis of the different

kinds of rule-based reasoning that can be modelled. In our view, applying an AGM-like approach on top of input/output systems allows an essential characterisation of change to be developed that focuses here on normative reasoning, but can actually be extended to other forms of rule-based reasoning.

2 Formal Framework

In this section, we briefly introduce the formal framework we will adopt in our analysis of normative change. In the last few decades, the area of knowledge representation and reasoning has proposed various formal approaches to modelling the dynamics of knowledge, and to modelling normative change in particular. As a result, one methodological issue that we need to address is what kind of analysis do we want to develop for normative change.

2.1 The AGM Approach

We will rely on the methodology of the *AGM approach* to belief change that we introduced in section 1.4. In the last 30 years, AGM has been the most popular formal approach to analysing the dynamics of beliefs, but it has been debated whether it is the best approach to analysing belief change in general, and normative change in particular. In this section, we briefly outline the main characteristics of this approach for the unfamiliar reader, and discuss why we still consider it to be a viable option for analysing normative change.

Let's start with a well-known example. Our knowledge base contains the following information:

- a. Sweden is an European country.
- b. All European swans are white.
- c. The bird I just caught in the trap is a swan.
- d. The bird I just caught in the trap is from Sweden.
- e. No bird can be black and white at the same time.

This information entails that the bird I just caught in a trap is white. But then I look at it and I see that it is undoubtedly black. I add to my knowledge base the following proposition:

- f. The bird I just caught in the trap is black.

From my knowledge base, I must conclude that the bird I just caught in the trap is both white and black. My knowledge base contains conflicting information, it is inconsistent. How should the situation be fixed? What constraints should we follow in changing our beliefs? And how should we give a formal characterisation to such constraints?

It is generally assumed that the constraints that a *rational* form of belief change should respect are based on considerations of two kinds:

1. *Logic*. Here the focus is on *consistency preservation*: the content of our knowledge base should always be devoid of contradictions.

Looking at our example, we cannot accept that we can believe that a bird is black and that the bird is white at the same time. Once we rely on piece of information f , we need to change the content of our knowledge base, since propositions a - f together necessarily imply a contradiction.

2. *Pragmatic*. This point and Point 1 above are intertwined. If we are forced to modify the content of our knowledge base in order to satisfy logical constraints, e.g. in order to preserve consistency, we should do so taking into consideration also pragmatic issues, based on, for example, *economy of information*. According to that principle, information is valuable, some pieces of information are more relevant and reliable than others, and if we are forced to drop some pieces of information, we should “minimise the damage” by eliminating only the minimal amount of information that is necessary to preserve logical consistency.

What should we do in our example once we learn proposition f and we spot the conflict? We could simply erase the entire knowledge base, just eliminate all the propositions (a)-(e). But why should we do this given that, for example, it is sufficient to drop only one proposition among (a), (b), (c), (d), and (e)?

In order to describe belief change, the AGM approach gives a formal definition to the knowledge representation desiderata by defining formal constraints based on logical or pragmatic considerations.

To formally introduce the AGM approach, we need some formal preliminaries. We use a classical propositional language \mathcal{L} , built from atomic propositional letters and using the propositional connectives $\neg, \wedge, \vee, \rightarrow, \equiv, \perp$. Lower-case letters a, b, c, \dots, x, y, z will be used to represent propositions. A *knowledge base* is a set of propositional formulas, that will be indicated by capital letters as \mathcal{K} . In addition, \models and Cn will represent the classical propositional entailment relation and entailment operator respectively.

The epistemic status of an agent is characterised by a knowledge base \mathcal{K} . Actually, the classical AGM approach embraces a perspective that has been dominant in epistemic logics: the epistemic status of the agent is characterised using a *belief set*, a logical theory closed under Cn . That is, the epistemic status of an agent is characterised by a knowledge base \mathcal{K} such that $\mathcal{K} = Cn(\mathcal{K})$. Let \mathcal{T} be the set of the belief sets (i.e. the closed theories) of language \mathcal{L} , that is, $\mathcal{T} := \{\mathcal{K} \subseteq 2^{\mathcal{L}} \mid \mathcal{K} = Cn(\mathcal{K})\}$.

The first question we need to address is what kind of changes we should consider. The AGM approach recognises three operations as the basic ones: *expansion*, *contraction*, and *revision*. Assume our agent A has a knowledge base \mathcal{K} :

- *Expansion* $+$: A is informed that proposition p holds, and simply adds it to \mathcal{K} without caring whether this could generate some contradiction. The resulting knowledge base is indicated as $\mathcal{K} + p$.
- *Contraction* $-$: A believes that p holds ($p \in \mathcal{K}$), but then decides that it is better to abandon such a belief, for example because the source is not considered trustworthy anymore. The resulting knowledge base, indicated as $\mathcal{K} - p$, should be such that p is no longer implied by A 's knowledge base.
- *Revision* $*$: A is informed that proposition p holds, and wants to add it to \mathcal{K} , but with the proviso that the resulting knowledge base should be logically sound. The resulting knowledge base is indicated as $\mathcal{K} * p$.

These three kinds of operations can be characterised using the function

$$\bullet : \mathcal{T} \times \mathcal{L} \mapsto \mathcal{T} \text{ with } \bullet \in \{+, -, *\}.$$

Actually, the truly basic operations are generally considered to be the first two, *expansion* and *contraction*, since *revision* is usually built on top of those using the so-called *Levi Identity* [40]:

$$\mathcal{K} * p := (\mathcal{K} - \neg p) + p.$$

Revising knowledge base \mathcal{K} by introducing a new proposition p requires that we guarantee that there are no pieces of information in our knowledge base that are in conflict with p . The reasonable way of obtaining this is to contract \mathcal{K} to ensure that it does not imply $\neg p$, and only then introduce p . This is the revision procedure that is modelled by the Levi Identity.

In the swan example, in order to revise the belief set with the information that the swan is black, we should proceed as follows: the belief set corresponds to the set $\mathcal{K} := Cn(\{a, b, c, d, e\})$ and we want to introduce f (“The swan in the trap is black”). Using the Levi Identity, the revision

$$\mathcal{K} * f$$

will consist in first contracting the piece of information $\neg f$ (“It is not the case that the swan in the trap is black”) from \mathcal{K} . The resulting belief set, $\mathcal{K} - \neg f$, should be a set of formulas that is smaller than \mathcal{K} and does not imply $\neg f$ anymore. For example, let us opt for weakening proposition b (“All European swans are white”) into a new proposition b' (“All European swans are white or black”), that is, $\mathcal{K} - \neg f = Cn(\{a, b', c, d, e\})$, and it is easy to check that $\mathcal{K} - \neg f$ does not imply $\neg f$ anymore. Only after the contraction do we add f , that is, we can set $\mathcal{K} * f = (\mathcal{K} - \neg f) + f = Cn(Cn(\{a, b', c, d, e\}) \cup \{f\})$, that is, $\mathcal{K} * f = Cn(\{a, b', c, d, e, f\})$.

We also have a complementary construction, the *Harper Identity*, in which revision is the primitive operator and contraction is defined on top of it:

$$\mathcal{K} - p := (\mathcal{K} * \neg p) \cap \mathcal{K}.$$

$\mathcal{K} - p$ should be a subset of \mathcal{K} not implying p , while $\mathcal{K} * \neg p$ should be a theory as close as possible to \mathcal{K} that implies $\neg p$ and does not imply p . The meaning of the Harper Identity is that since $\mathcal{K} * \neg p$ should not imply p , if we intersect it with \mathcal{K} , we obtain a contraction: a subset of \mathcal{K} that does not imply p .

We can rephrase the above example to show that the Harper Identity and the Levi Identity can correspond to each other. Let \mathcal{K} be our knowledge base containing propositions (a)-(e), and assume that we have a revision operator $*$, as described above and which is introduced here as a primary operator, such that $\mathcal{K} * f = Cn(\{a, b', c, d, e, f\})$. If we use the Harper Identity to define a contraction operator $-$ from $*$, we obtain $\mathcal{K} - f = Cn(\{a, b', c, d, e, f\}) \cap Cn(\{a, b, c, d, e\})$ that, since $b \equiv b'$, corresponds to $\mathcal{K} - f = Cn(\{a, b', c, d, e\})$, that is, the contraction we have used above as a primitive operator to define $*$ via the Levi Identity. In what follows, we will use both Levi and Harper Identities, and we will soon give a more formal definition of the correspondence between the two.

Once we have identified the basic operations we are interested in, the second question we need to address is how we want to model and constrain such change operations. For each kind of operation, we want to determine a set of desired properties they should satisfy, and give a formal expression to such desiderata.

Expansion is considered to be a trivial operation, formalised by adding the formula we are interested in to the knowledge base and letting the agent commit to all the logical consequences of such an addition:

$$\mathcal{K} + a := Cn(\mathcal{K} \cup \{a\}).$$

In the *contraction* operation, an agent starts with a belief set \mathcal{K} (e.g. the theory determined by sentences (a)-(e) above) and wants to eliminate some pieces of information in the belief set (e.g. that the swan is white). The AGM approach gives a formal representation to a basic set of desiderata using six *postulates*.

Definition 6 (AGM contraction [5]). *Let $-$ be a function that, given a belief set \mathcal{K} and a proposition a , returns a new belief set $\mathcal{K} - a$. Function $-$ is an AGM basic contraction operator iff it satisfies the following postulates:*

- (- 1) $\mathcal{K} - a$ is closed under Cn (closure)
- (- 2) $\mathcal{K} - a \subseteq \mathcal{K}$ (inclusion)
- (- 3) If $a \notin \mathcal{K}$, then $\mathcal{K} - a = \mathcal{K}$ (vacuity)
- (- 4) If $\not\models a$, then $a \notin \mathcal{K} - a$ (success)
- (- 5) If $a \in \mathcal{K}$, then $\mathcal{K} \subseteq (\mathcal{K} - a) + a$ (recovery)
- (- 6) If $\models a \equiv b$, then $\mathcal{K} - a = \mathcal{K} - b$ (extensionality)

Two extra postulates are introduced to relate the contraction of complex formulas to the contraction of their components:

- (- 7) $\mathcal{K} - a \cap \mathcal{K} - b \subseteq \mathcal{K} - (a \wedge b)$ (conjunctive overlap)
- (- 8) If $a \notin \mathcal{K} - (a \wedge b)$, then $\mathcal{K} - (a \wedge b) \subseteq \mathcal{K} - a$ (conjunctive inclusion)

Function $-$ is an AGM contraction operator iff it satisfies postulates (- 1)-(- 8).

We will briefly go through the meaning of these postulates. Postulate (- 1) enforces an idealisation we have already discussed: the epistemic status of the agent is described using logically closed theories (belief sets), hence every change operation must transform a closed theory into a new closed theory. Postulate (- 2) imposes that the change operation must result in an actual *contraction* of the agent's belief set, that is, the set of formulas believed by the agent at the end is a subset of the initial beliefs. Postulate (- 3) formalises a principle of an economical nature: if the contraction operation involves a formula that is already excluded from the agent's beliefs, the contraction operation is *vacuous*, that is, nothing changes, since the desired result is already satisfied. Postulate (- 4) imposes that, whenever possible, that is, whenever the formula to be contracted is a *contingent* formula and not a

tautology, the contraction operation must be successful, that is, the formula should no longer be in the resulting belief set. Let us jump to postulate (– 6), leaving postulate (– 5) aside for one moment. Postulate (– 6) imposes independence from syntax, which is a classical logical principle: whenever two pieces of information are logically equivalent, they are indifferent from a logical point of view, and their impact on the agent’s belief set is exactly the same. It is easy to see that this principle is strongly related to postulate (– 1), the use of logically closed theories to model the epistemic states. While the use of closed theories imposes indifference with regard to the syntactic form of the knowledge base in the *static* model of the agent’s epistemic state, the principle of *extensionality* extends such syntactic indifference also to operations modelling the *dynamics* of the agent’s epistemic states. Postulates (– 7) and (– 8) are considered extra postulates, since they are the only ones that impose constraints on the way a contraction operator behaves with different formulas, in particular how the contraction of a formula should behave with the contraction of logically weaker formulas.

Postulate (– 5), *recovery*, has a special status, since, probably together with postulate (– 1), it is the most debated AGM principle, and in a certain sense it is also the one that mainly characterises the classical AGM approach. Its nature is purely economical, based on the idea that in order to contract, we “cut” as little as possible from the original knowledge base. So little that if the agent decides that contracting by formula a was not a good idea and that a should be added back, we should be able to return to the original knowledge base without any loss. In fact, together with postulate (– 2), postulate (– 5) implies that if $a \in \mathcal{K}$, then $\mathcal{K} = (\mathcal{K} - a) + a$, that is, if we put a back after a contraction, we go back to the initial state. It has been debated extensively whether *recovery* is a reasonable principle for contraction, and we will return to this issue later in this section.

Anyway, the reader can see that each of these eight postulates answers to either logical or pragmatic desiderata. For a more detailed explanation of their meaning, we refer the interested reader to the original AGM paper [5] and many other publications in the field.

It is worth mentioning that Rott [67] has disputed whether the AGM approach does actually satisfy any principle of informational economy. Despite the relevance of Rott’s observations, postulates like (– 3) and (– 5) are generally seen as necessary conditions for defining contraction operators that satisfy the principle of informational economy. The principle of informational economy, which has been expressed in various forms and with different names, has always been addressed by researchers in the area as the main guideline for the definition of postulates.

In our presentation of AGM belief change, we first introduced a set of possible change operations (specifically, *expansion*, *contraction*, and *revision*), and then a set

of *postulates* to give formal expression to the properties we think such operations should satisfy, specifically those for contraction. The next step is to present the formal tools that we can use to define such change operators. That is, given a set of postulates, the AGM approach is focused on providing a formal characterisation of the class of operations that satisfy such postulates. The classical results in the area define classes of change operations using maxiconsistent subsets and choice functions [5], orderings over possible-world semantics representing which situations the agent considers to be more plausible [33, 37], or orderings over the formulas (*epistemic entrenchment relations*) indicating which pieces of information the agent considers to be more or less reliable [27].

Regarding contraction, the initial characterisation of the class of operations satisfying the basic postulates is based on identifying the maximal subsets of the belief set that do not imply the contracted formula. The resulting belief set is defined by the intersection of some such maximal subsets. Which maximal subsets are used in the definition of the contraction is formalised via a dedicated choice function.

Definition 7 (Partial meet contraction [5, p. 512]).

Let $\mathcal{K} \perp a$ be the remainder set, containing the maximal subsets \mathcal{K}' of \mathcal{K} such that \mathcal{K}' is a closed theory and $a \notin \mathcal{K}'$. That is, $\mathcal{K}' \in \mathcal{K} \perp a$ iff

- (i) $\mathcal{K}' \subseteq \mathcal{K}$,
- (ii) $\mathcal{K}' \in \mathcal{T}$,
- (iii) $a \notin \mathcal{K}'$, and
- (iv) there is no set $\mathcal{K}'' \in \mathcal{T}$ such that $\mathcal{K}' \subset \mathcal{K}'' \subseteq \mathcal{K}$ and $a \notin \mathcal{K}''$.

Let pm be a choice function defined over the set of the remainder sets. Function pm is a partial meet function if for every KB \mathcal{K} and every formula a :

- $pm(\mathcal{K} \perp a) \subseteq \mathcal{K} \perp a$, and
- if $\mathcal{K} \perp a \neq \emptyset$, then $pm(\mathcal{K} \perp a) \neq \emptyset$.

A partial meet contraction operator $-$ is defined as: $\mathcal{K}_A^- = \bigcap pm(\mathcal{K} \perp A)$.

The class of partial meet contractions is sufficient to give an operational characterisation of the class of AGM basic contraction operations.

Observation 8. [5, Observation 2.5] A contraction operator $- : \mathcal{T} \times \mathcal{L} \mapsto \mathcal{T}$ is an AGM basic contraction operator (satisfying (– 1)-(– 6)) iff it is a partial meet contraction operator.

An analogous analysis can be developed for *revision*. First of all, we can formalise our desiderata via appropriate postulates.

Definition 9 (AGM revision $*$ [5]). *Let $*$ be a function that, given a belief set \mathcal{K} and a proposition a , returns a new belief set $\mathcal{K} * a$. Function $*$ is an AGM basic revision operator iff it satisfies the following postulates:*

- (* 1) $\mathcal{K} * a$ is closed under Cn (closure)
- (* 2) $a \in \mathcal{K} * a$ (success)
- (* 3) $\mathcal{K} * a \subseteq \mathcal{K} + a$ (inclusion)
- (* 4) If $\neg a \notin \mathcal{K}$, then $\mathcal{K} + a = \mathcal{K} * a$ (vacuity)
- (* 5) $\perp \in (\mathcal{K} * a)$ iff $\models \neg a$ (triviality)
- (* 6) If $\models a \equiv b$, then $\mathcal{K} * a = \mathcal{K} * b$ (extensionality)

Two extra postulates are introduced also for revision. These postulates relate the revision of complex formulas to the revision of their components:

- (* 7) $\mathcal{K} * (a \wedge b) \subseteq (\mathcal{K} * a) + b$ (Iterated (* 3))
- (* 8) If $\neg b \notin \mathcal{K} * (a)$ then $(\mathcal{K} * a) + b \subseteq \mathcal{K} * (a \wedge b)$ (Iterated (* 4))

Function $*$ is an AGM revision operator iff it satisfies the postulates (* 1)-(* 8).

The meaning of the postulates for revision is very close to the meaning of the postulates for contraction. The parallel is clear for postulates (* 1), (* 2), (* 3), (* 4), (* 6) and the correspondent postulates for contraction. Postulate (* 5) imposes perhaps the key rational desideratum for modelling belief dynamics: preserving consistency. Whenever we add a new piece of information a , the only case where the resulting belief set can be inconsistent is when a itself is inconsistent.

We briefly summarise a series of well-known basic results in the area that show how the notions introduced up to this point are solidly connected to one other in AGM theory. First of all, the construction of AGM revision and contraction operators are intertwined via the Levi Identity.

Observation 10. [5] *Let $*$: $\mathcal{T} \times \mathcal{L} \mapsto \mathcal{T}$ be a revision operator. Function $*$ is a basic AGM revision operator (it satisfies (* 1)-(* 6)) if and only if there is a contraction operator – such that:*

- $*$ can be defined via the Levi Identity from $-$. That is, for every \mathcal{K} and a ,

$$\mathcal{K} * a = (\mathcal{K} - \neg a) + a$$

- $-$ is a basic AGM contraction operator (it satisfies $(- 1)$ - $(- 6)$).

Given Observation 8, Observation 10 connects the construction of basic AGM revision operators to the class of partial meet contractions via the Levi Identity.

An analogous result [5] holds for contraction and revision operators satisfying postulates $(- 1)$ - $(- 8)$ and $(* 1)$ - $(* 8)$ respectively.

Such a dependency of revision on contraction can also be reversed, moving from AGM revision operators to the definition of AGM contraction operators: the one-to-one correspondence between the Levi Identity and the Harper Identity, that we have briefly exemplified above in revising and contracting our knowledge base about swans, can actually be formally proved. Let us translate the Levi and Harper Identities into transformation functions. Given a belief set \mathcal{K} , a formula a , a contraction operator $-$ and a revision operator $*$, let

- $\mathcal{K} \mathbb{R}(-) a := Cn((\mathcal{K} - a) \cup \{a\})$
- $\mathcal{K} \mathbb{C}(*) a := (\mathcal{K} * \neg a) \cap \mathcal{K}$

where $\mathbb{R}(-)$ represents a revision operator obtained from contraction $-$ via the Levi Identity and $\mathbb{C}(*)$ represents a contraction operator obtained from revision $*$ via the Harper Identity. Using these operators, Makinson has proven that there is full correspondence between the Levi and Harper Identities.

Observation 11. [41] *Let \mathcal{K} be a belief set, and let a be a formula, with $\mathbb{R}(-)$ and $\mathbb{C}(*)$ defined as above.*

- *Let $-$ satisfy the postulates of closure, inclusion, vacuity, extensionality, and recovery. Then $\mathbb{C}(\mathbb{R}(-)) = -$.*
- *Let $*$ satisfy the postulates of closure, inclusion, success, and extensionality. Then $\mathbb{R}(\mathbb{C}(*)) = *$.*

As an immediate consequence, the Levi and Harper Identities have been shown to be interchangeable for AGM theory:

$$\mathcal{K} * a = (\mathcal{K} \cap \mathcal{K} * a) + a;$$

$$\mathcal{K} - a = \mathcal{K} \cap ((\mathcal{K} - a) + \neg a).$$

What we have presented up to this point are some key results of the AGM approach that provide an essential introduction to the unfamiliar reader, and which are relevant to the sections that follow.

2.2 Criticisms of the AGM Approach

Simplifying, we could say that there are three main steps that characterise the AGM method:

- the identification of the typologies of change we want to model and of the properties we want them to satisfy;
- the translation of such desiderata into postulates, that is, into formal constraints;
- the characterisation of the classes of operators that satisfy the desired set of postulates. Such a characterisation is usually obtained by proving the correspondence of such operators to a class of constructions defined using a relevant formal tool (e.g. maxiconsistent sets, possible-world models. . .).

The AGM approach to belief change has quickly become standard in the field, and the last 30 years has seen many contributions [25]. Despite the fact that it has become a major research topic in knowledge representation, it is an approach that has been frequently and heavily criticised, and new lines of research have sprouted from some of these critiques. We briefly list some of the main critiques the AGM approach has received.

2.2.1 Too Many Constraints Imposed on the Underlying Logic

The AGM approach was originally developed for classical propositional logic (PL), and the classical results assume that the underlying logic, characterised by a language L and an entailment operator Cn , satisfies many of the formal properties that characterise PL:

1. The language L is closed under the propositional connectives.
2. The entailment operator Cn is *Tarskian*, that is, given two sets of formulas $\mathcal{K}, \mathcal{K}' \subseteq L$, it satisfies the following properties:
 - *monotonicity*: if $\mathcal{K} \subseteq \mathcal{K}'$, then $Cn(\mathcal{K}) \subseteq Cn(\mathcal{K}')$;
 - *idempotence*: $Cn(\mathcal{K}) = Cn(Cn(\mathcal{K}))$;
 - *iteration*: $\mathcal{K} \subseteq Cn(\mathcal{K})$.
3. The consequence operator satisfies some well-known properties of classical logic:

- *deduction*: $b \in Cn(\mathcal{K} \cup \{a\})$ iff $(a \rightarrow b) \in Cn(\mathcal{K})$;
- *disjunction in the premises*: if $a \in Cn(\mathcal{K} \cup \{b\})$ and $a \in Cn(\mathcal{K} \cup \{c\})$, then $a \in Cn(\mathcal{K} \cup \{b \vee c\})$.

4. *Compactness*: if $a \in Cn(\mathcal{K})$, then $a \in Cn(\mathcal{K}')$ for some finite $\mathcal{K}' \subseteq \mathcal{K}$.

Much recent research in belief revision has been dedicated to investigating whether the above constraints are essential to the definition of AGM operators and, when we are dealing with an underlying logic that does not allow the definition of classical AGM postulates, what other meaningful postulates can be defined and satisfied. For example, the AGM approach has been applied to logics that are not fully closed under propositional operators [21, 80], that are not monotonic [79, 20, 19], and that are not compact [65].

This article will also deal with a family of logics that do not satisfy all the properties listed above. Input/output logics are not closed under propositional operators and, because of that, cannot satisfy properties like *deduction* and *disjunction in the premises*. Some input/output logics also do not satisfy the property of *monotonicity* [43]. Although we shall not discuss them in this article, application of the AGM methodology to normative change based on non-monotonic input/output logics is a promising field of inquiry.

2.2.2 Lack of Expressiveness

It has often been pointed out that the expansion/contraction/revision triad is not sufficient to account for the dynamics of information. It is also claimed that the AGM approach is not appropriate for handling multi-agent systems because it is suitable only for factual information.

With respect to the first line of criticism, it is worth mentioning that operations that are not reducible to the original ones have been introduced, such as *update* [36] and *merging* [39] among others. Besides, many refinements to the original operations have been proposed, based on alternative postulates and formal constructions, which introduce new dimensions to the original operations, such as the trustworthiness of the new information [25, Chapter 8]. Despite being a common place that the AGM operations of contraction and revision are not sufficient to cover all the relevant dynamics of information, it is generally accepted that analysing the operations of contraction and revision is a good starting point for modelling informational change in many contexts. Analysing contraction and revision in different formal contexts allows us to deal with the ideas of minimal change and consistency preservation in

each of those contexts, and minimal change and consistency preservation are the two main stepping stones towards characterising rational informational change.

It is true that multi-agent contexts are not immediately compatible with the AGM approach, since some classical AGM postulates would be counter-intuitive in such a framework.

In the area of Dynamic Epistemic Logic (DEL), it has been pointed out that some sentences, for example those resembling the structure of that used in *Moore's paradox*, are not compatible with the *success* postulate [76]. The DEL framework allows us to model the dynamics of epistemic states in which the agent also models higher-order sentences representing beliefs about its own beliefs and the beliefs of other agents. On the other hand, AGM is easier to understand, and allows a more in-depth analysis of specific kinds of operations. Working first at the AGM level, and later transporting the proposed solutions to other frameworks such as the DEL framework, can be seen as a good research strategy. Also, some domains, like formal ontologies or the domain under consideration in this article, normative bodies, do not usually need to deal with a multi-agent aspect in modelling change.

2.2.3 Logical Closure and the Recovery Postulate

Finally, let us consider two further lines of criticisms of the AGM approach that are particularly relevant for what follows. These are connected to the *recovery* (-5) and the *closure* ($(-1)/(*1)$) postulates.

As mentioned above in this section, the recovery postulate has often been criticised. On the one hand, its desirability is intertwined with the use of logically closed belief sets. On the other hand, as many commentators have pointed out, the recovery postulate is not always desirable even if we are working with closed belief sets (see [25, Sect. 5.1] for an overview).

Moreover, if we define revision on top of contraction via the Levi Identity, it turns out that the recovery postulate is not necessary to characterise the class of the AGM basic revision operators. That is, the representation that results in Observation 10 remains valid if we drop postulate (-5).

Observation 12. [28] *Let $* : \mathcal{T} \times \mathcal{L} \mapsto \mathcal{T}$ be a revision operator. Function $*$ is a basic AGM revision operator (it satisfies $(*1)$ - $(*6)$) if and only if there is a contraction operator $-$ such that:*

\cdot $$ can be defined via the Levi Identity from $-$. That is, for every \mathcal{K} and a ,*

$$\mathcal{K} * a = (\mathcal{K} - \neg a) + a.$$

· – satisfies (– 1)-(– 4) and (– 6).

The criticisms of the recovery postulate, together with the fact that it is not a necessary property in order to characterise well-behaved revision operators, has convinced many researchers to drop such a postulate in many contexts, looking for more significant alternatives [24].

As mentioned above, the AGM approach models change over belief sets, that is, it does not consider arbitrary sets of formulas, but only logically closed theories.

This is a constraint that is in line with the classical modelling approach of epistemic logics, and it is prone to the same kind of criticisms. On the one hand, characterising epistemic states as closed logical theories is seen as the correct way to characterise rational agents, since it allows a description of knowledge that is syntax-independent and that models the commitment a rational agent should have towards all the consequences of what is explicitly stated in a knowledge base. On the other hand, depending on the modelling goals, exactly the same arguments can be considered as drawbacks. If we investigate the belief states and dynamics of agents with bounded rationality, committing to closed logical theories is too strong an idealisation, which in epistemic logics is labelled as *logical omniscience*. Moreover, the syntactic form of the knowledge base can actually play a role in modelling the way the agent manages the information at its own disposal, for example by making explicit how the agent clusters pieces of information together in a single formula. The belief change community has reacted by developing the theory of *base revision*, where the same approach as AGM to investigation is applied to finite knowledge bases rather than logically closed theories [35].

2.3 Base Contraction and Revision

In base revision, the epistemic status of an agent is described using a set of formulas K that is not necessarily logically closed. The basic operation in base revision is Hansson’s *kernel contraction* [35], which is a re-interpretation at the level of finite base of the AGM notion of contraction based on remainder sets.

Hansson’s base contraction is based on the notions of *kernels* and *incision functions* in a way that resembles the roles of the *remainder sets* and the *partial meet functions* in partial meet contraction. Given a knowledge base K and a formula a , the a -*kernels* of K are the minimal subsets of K that have a as a logical consequence. Eliminating some pieces of information from each kernel allows us to avoid deriving a , and such an elimination is made using an *incision function*.

Definition 13 (Kernel set and incision function [35]). *Let $a \in \mathcal{L}$ and $K \subseteq \mathcal{L}$. The set $\text{Kern}_K(a) \subseteq 2^{2^{\mathcal{L}}}$ is the kernel set of K with respect to a if it is defined as follows. $X \in \text{Kern}_K(a)$ if and only if:*

- $X \subseteq K$;
- $a \in \text{Cn}(X)$;
- if $X' \subset X$, then $a \notin \text{Cn}(X')$.

An incision function σ defined over the kernel sets is a choice function such that:

- $\sigma(\text{Kern}_K(a)) \subseteq \bigcup \text{Kern}_K(a)$;
- $\sigma(\text{Kern}_K(a)) \cap X \neq \emptyset$ for all $X \in \text{Kern}_K(a)$.

Once the incision function has specified the information that should be eliminated from K in order to avoid deriving a , we can use it to define a contraction operator on arbitrary sets of formulas.

Definition 14 (Kernel contraction [35]). *Let $a \in \mathcal{L}$ and $K \subseteq \mathcal{L}$. Operator $-_{\sigma} : 2^{\mathcal{L}} \times \mathcal{L} \mapsto 2^{\mathcal{L}}$ is a kernel contraction operator if*

$$K -_{\sigma} a = K \setminus \sigma(\text{Kern}_K(a)).$$

Hansson gives a postulate characterisation of kernel contractions.

Observation 15. [35] *A function $- : 2^{\mathcal{L}} \times \mathcal{L} \mapsto 2^{\mathcal{L}}$ is a kernel contraction if and only if it satisfies the following postulates:*

- ($-_{\sigma}$ 1) $K - a \subseteq K$ (inclusion)
- ($-_{\sigma}$ 2) If $\nexists a$, then $a \notin K - a$ (success)
- ($-_{\sigma}$ 3) If $b \in K \setminus K - a$, then there is a $K' \subset K$ such that $a \notin \text{Cn}(K')$ but $a \in \text{Cn}(K' \cup \{b\})$ (core-retainment)
- ($-_{\sigma}$ 4) If for all subsets K' of K , it holds that $a \in \text{Cn}(K')$ iff $b \in \text{Cn}(K')$, then $K - a = K - b$ (uniformity)

We can also define revision combining contraction and expansion using bases, but now we have two possible ways of combining the two operations [34],

- $\mathcal{K} *__{\sigma} a = (\mathcal{K} -_{\sigma} \neg a) +_{\sigma} a$ (Levi Identity)
- $\mathcal{K} *__{\sigma} a = (\mathcal{K} +_{\sigma} a) -_{\sigma} \neg a$ (Reversed Levi Identity)

where $\mathcal{K} +_{\sigma} a := \mathcal{K} \cup \{a\}$. The two options define revision operators with different properties [34]. The Reversed Levi Identity is not a viable option when we are working with belief sets, since the first step, the expansion, could take us to an inconsistent theory, the contraction of which is not efficiently managed by the classical AGM approach.

3 Formal Analysis of Normative Change

The distinction between norms and obligations was articulated and formally developed in input/output logic [42]. Input/output logic takes a very general view of the process used to obtain conclusions (outputs) from given sets of premises (inputs). To detach an obligation from a norm, there must be a context, and the norm must be conditional. Thus, norms are just particular kinds of rules, and one may view a normative system simply as a set of rules.

Makinson's iterative approach to normative reasoning distinguishes unconstrained from constrained output. Unconstrained is close to classical logic, whereas constrained output is much less similar, due to the existence of multiple output sets (or extensions), for example. Examples of constrained output are default reasoning, defeasible deontic reasoning etc.

Makinson and van der Torre introduced seven distinct input/output logics, including both a semantic definition and a proof theoretic characterisation [43, 44]. They showed that their seven unconstrained input/output logics cannot handle contrary-to-duty reasoning and thus cannot be used as logics representing normative reasoning. They therefore introduced constrained output in a companion paper, and they showed how that can be used as a logic of norms. However, the user has to make some seemingly arbitrary choices by, for example, choosing between a sceptical and a credulous approach. Moreover, the complex nature of constrained output makes it difficult to handle. This becomes apparent if we consider norm change, like contraction and revision of norms. The constrained input/output logic framework becomes relatively complex and cumbersome. Here, we follow the work of Boella *et al.* [16] and call the generators of unconstrained output *rules*.

3.1 Input/Output Logic

In this section, we give a general introduction to input/output logic. For a deeper look into the input/output logic framework, the reader is referred to the work of Makinson and van der Torre [45] and Parent and van der Torre [57].

A *rule* is a pair of propositional formulas,⁶ called the antecedent and consequent of the rule.

Definition 16 (Rules [42]). *Let L be a propositional logic built on a finite set of propositional atoms A . A rule-based system $R \subseteq L \times L$ is a set of pairs of L , written as $R = \{(a_1, x_1), (a_2, x_2), \dots, (a_n, x_n)\}$.*

Rules allow the derivation of formulas, like the derivation of obligations and prohibitions in a legal code. Which obligations and prohibitions can be derived depends on the factual situation (i.e. the *context* or *input*), which is a propositional formula.

Definition 17 (Operational semantics [42]). *An input/output operation $out : \mathcal{P}(L \times L) \times L \rightarrow \mathcal{P}(L)$ is a function from the set of rule-based systems and contexts to a set of sentences of L .*

Note that operator *out* satisfies the principle of irrelevance of syntax. The simplest input/output logic defined by Makinson and van der Torre is the so-called simple-minded output.

Definition 18 (Simple-minded output [42]). *Proposition x is in the simple-minded output of the set of rules R in context a , written as $x \in out_1(R, a)$, if there is a set of rules $(a_1, x_1), \dots, (a_n, x_n) \in R$ such that $a_i \in Cn(a)$ and $x \in Cn(x_1 \wedge \dots \wedge x_n)$, where $Cn(a)$ is the consequence set of a in L .*

A set of rules is said to ‘imply’ another rule (a, x) if and only if x is in the output in context a .

Definition 19. *Rule ‘implication’ by Makinson and van der Torre [42]] Rule (a, x) is ‘implied’ by rule-based system R , written as $(a, x) \in out(R)$, if and only if $x \in out(R, a)$.*

As Makinson and van der Torre observe, the relation between the ‘implication’ among rules $(a, x) \in out(R)$ and the ‘operational semantics’ $x \in out(R, a)$ has an analogy in classical logic, where the pair $a \models x$ is equivalent to the membership of x in the consequence set of a , written as $x \in Cn(a)$.

Definition 20. [16] *Function out is a closure operation when the following three conditions hold:*

⁶One may also use a first-order, temporal or action logic. The choice of classical propositional logic is intended to stay closer to the AGM theory.

reflexivity: $x \in out(R \cup \{(a, x)\}, a)$ (in other words, $R \subseteq out(R)$), and if the context is precisely the antecedent of one of the rules, then the output contains the consequent of that rule.

monotony: $x \in out(R_1, a)$ implies $x \in out(R_1 \cup R_2, a)$ (in other words, $out(R_1) \subseteq out(R_1 \cup R_2)$), and if the set of rules increases, then no conclusions are lost.

idempotence: if $x \in out(R, a)$, then for all b , we have $out(R, b) = out(R \cup \{(a, x)\}, b)$ (in other words, $out(R) = out(out((R)))$), and if x is obligatory in context a , then (a, x) can be added to the rule-based system without changing the output.

Makinson and van der Torre show that their seven input/output logics satisfy the Tarskian properties, and their notion of ‘implication’ among rules is therefore a Tarskian consequence relation, a crucial characteristic to incorporating the AGM construction into the framework of input/output logics.

Definition 21. [42] Let $R(a) = \{x \mid (a, x) \in R\}$, and let v be a classical valuation (maxiconsistent set of propositions) or L . Simple-minded, basic, reusable and basic reusable output are defined as follows:

simple minded: $out_1(R, a) = Cn(R(Cn(a)))$

basic: $out_2(R, a) = \cap\{out_1(R, v) \mid a \in v\}$

reusable: $out_3(R, a) = \cap\{out_1(R, b) \mid a \in Cn(b), out_1(R, b) \subseteq Cn(b)\}$

basic reusable: $out_4(R, a) = \cap\{out_1(R, v) : a \in v \text{ and } out_1(R, v) \subseteq v\}$

Basic output handles reasoning by cases, and reusable output handles iterated detachment [42]. Moreover, for each input/output logic, a corresponding throughput operator is defined by:

$$out_i^+(R, a) = out_i(R \cup \{(b, b) \mid b \in L\}, a).$$

As many of the examples discussed in section 1 have shown, normative change has to handle and solve inconsistencies and incoherencies (on the concept of incoherence, see section 3.2 below) between obligations and permissions as two distinctive kinds of regulative rules.

The implication (or derivation) of obligations from a set O of obligatory regulative rules is given by definition 19. With respect to permissions, it is important

beforehand to distinguish, following Alchourrón [2], between *weak* (or negative) permissions and *strong* (or positive) permissions. In its weak sense, a permission to x in context b is just the absence of a prohibition to x in context b . That is, if we consider a set of obligation rules O , then a permission $\langle a, x \rangle$ is implied by O if and only if $\neg x \notin \text{out}(O, b)$ [44].

In its strong or positive sense, a permission is derived from explicit enactments of obligations as well as permissive rules. The output of a set of explicit permissions is defined below:

Definition 22. [44] *Let O be a set of obligations and let $P \subseteq (L \times L)$ be a set of explicit permissions. Then, $(a, x) \in \text{perm}_i(P, N)$ iff $(a, x) \in \text{out}_i(O \cup Q)$ for some singleton or empty $Q \subseteq P$.*

As we have emphasised in section 1.1 when referring to the problem called *network effects*, some difficulties concerning normative change are related to the combination of constitutive and regulative rules in the normative system.

We may model this problem using input/output logics by making the output of a normative set (possibly joined with the input set) the input of the output operation on the other normative set. It is also possible to combine sets for deriving obligations and explicit permissions.

A typical combination of normative sets is given by the definition or qualification, by a constitutive rule, of a concept present in a regulative rule. For instance, a data protection legislation contains a regulative rule establishing that consent by the data subject (*consent*) is a condition for lawful processing of his/her personal data (*process*). Suppose that a platform processes the personal data of its users without explicit consent, considering that authorisation is implicit unless they explicitly object to that processing (*opt-out* model). If an user of that internet platform has not opted out, would the processing of her personal data be lawful? The answer may be found in a constitutive rule stipulating that only the data user's explicit and written authorisation for processing counts as consent (*opt-in* model). If the set of constitutive rules contain such a rule, then an opt-out model does not count as valid consent for personal data processing. This example of legal reasoning may be modelled by a combination of a set C of constitutive rules and sets O and P of regulative rules, where an output operator on the set of constitutive rules delivers the inputs for the output operator on the sets of regulative rules.

We shall use a general definition of the relation between a constitutive and a regulative rule in a derivation:

Definition 23. *Let $A \subseteq L$, $I \in \{A, \emptyset\}$, let out_i and out_j be output operators, and let C and R be constitutive and regulative sets of rules respectively. Then, the combined output of C and R is defined as:*

$$out_{i,j}(C, R, A) = out_i(R, out_j(C, A) \cup I).$$

The definition and the results regarding the contraction operator in section 3.7 covers, with straightforward adaptations, both cases of combinations, i.e. constitutive with permissive rules and constitutive with obligation rules, as follows:

$$\begin{aligned} out_{i,j}(O, C, A) &= out_i(O, out_j(C, A) \cup I); \\ perm_{i,j}(P, C, A) &= perm_i(P, out_j(C, A) \cup I). \end{aligned}$$

In the examples used throughout this article, we shall consider combined *out* and *perm* operators in which $I = A$. To formalise the above example on consent for lawful data processing using a combination of sets of normative rules, let us consider the following normative sets:

$$\begin{aligned} C &= \{(opt-in, consent), (opt-out, \neg consent)\} \\ P &= \{(consent, process)\} \\ O &= \{(\neg consent, \neg process), \} \end{aligned}$$

The normative system implies that $(opt-in, process) \in perm_{1,1}(P, C)$ and that $(opt-out, \neg process) \in perm_{1,1}(O, C)$. That is, it is permitted to process personal data if authorisation was obtained by an opt-in model, while it is forbidden to process that data if the model used was opt-out.

3.2 Consistency and Coherence of Normative Systems

As example 4 in section 1.2 shows, constitutive rules may be responsible for genuine normative conflicts when combined with a regulative set. In order to model this feature, it should be possible to verify regulative sets that are consistent but whose combination with a constitutive set implies inconsistent conditional norms. To avoid confusion, let us qualify regulative sets as consistent or inconsistent and combinations of constitutive sets with regulative sets as coherent or incoherent.

Consistency is defined with respect to a given context. We say that a normative set N is *b-consistent* if and only if $(b, \perp) \notin out(N)$. Accordingly, a combination (C, R) is *b-coherent* if and only if $(b, \perp) \notin out(C, R)$. If we have a set of obligations O and a set of explicit permissions P , then such normative sets are *b-consistent* if and only if for any sentence x , it is not the case that $(b, x) \in perm(O, P)$ and $(b, \neg x) \in out(O)$. Accordingly, a combination of a set of constitutive rules and a set of obligations and the same set of constitutive rules and a set of permissions

is *b-coherent* if and only if, for any sentence x , it is not the case that $(b, x) \in perm(O, P, C)$ and $(b, \neg x) \in out(O, C)$.

When should we then consider a normative system to be generally consistent or coherent? We may consider two extreme possibilities for such definitions.

The first extreme would be to consider a normative system *consistent* if it is consistent for all possible inputs, that is, to demand \perp -consistency. This conception would limit the possibility of giving opposite commands in logically independent conditions, since $N = \{(a, x), (b, \neg x)\}$ would be inconsistent.

The other extreme would be to consider a normative system *consistent* if it is consistent for a tautological input, i.e. to demand \top -consistency. This conception also seems inadequate because normative sets with genuine conflicts such as $N = \{(a, x), (a, \neg x)\}$ would be rendered consistent.

As a middle ground, we shall consider a normative set N consistent if it is *b-consistent* for every b such that $b \in Cl(a)$ and $a \in body(N)$ where $body(N) = \{b : (b, x) \in N\}$. That is, a normative set is consistent if there is no condition explicitly mentioned in its conditional rules that would, as input, deliver inconsistent outputs. Accordingly, a combination (C, R) is coherent if it is *b-coherent* for every b such that $b \in Cl(a)$ and $a \in body(C)$.

Therefore, we may have a consistent set R but an incoherent combination (C, R) , which would demand a contraction to restore coherence.

Let us formalise example 4 in the model proposed here. Following Maranhão and de Souza [52], we shall employ a basic reusable output operator (out_4) for the set of constitutive rules, and a basic output operator (out_2) for the sets of regulative (obligatory and permissive) rules. Recall that the example referred to a normative system where the police have the power to access (*acc*) property items (*prop*) in a search & seizure order (*sord*) but are forbidden from accessing ongoing communication (*com*) without an interception order (*iord*). The pertinent question is whether an exchange of messages stored on a mobile phone (*sms*) counts as data (*dat*) or as communication (or both). This normative system could be represented by the following normative sets of constitutive (C), regulative obligation (O) and regulative permission (P) rules:

$$\begin{aligned} C &= \{(sms, com), (sms, dat), (dat, prop)\} \\ P &= \{(prop \wedge sord, acc), (com \wedge iord, acc)\} \\ O &= \{(com \wedge \neg iord, \neg acc), (prop \wedge \neg sord, \neg acc)\} \end{aligned}$$

The corresponding normative theory is both consistent and coherent as there is no explicit condition in these normative sets that can, by itself, deliver a contradiction as output. However, given that a message exchange on a mobile phone

collected during an authorised search is both stored data and a form of ongoing communication, a search & seizure order to check message exchanges would deliver a contradiction, that is, we have both $(sms \wedge sord \wedge \neg iord, acc) \in perm_{2,4}(O, P, C)$ and $(sms \wedge sord \wedge \neg iord, \neg acc) \in out_{2,4}(O, C)$. Hence, the normative system is $(sms \wedge sord \wedge \neg iord)$ -incoherent, and a contraction should take place to restore coherence for that specific context.

There are different ways to reach this goal. And the task of legal interpretation, doctrinal or judicial, is to choose and justify such choices. It is possible to restore coherence by handling the definitions involved, that is, by contracting the set of constitutive rules, by contracting the set of regulative rules, or by deleting rules from both sets. We shall explore these alternatives in section 3.7 below.

3.3 Contraction of Normative Systems

Boella *et al.* [16] defined a rule set as a set of rules closed under an input/output logic ($out(R)$), and generalised the AGM postulates as postulates for the revision of norms. In order to keep an abstract approach and obtain general results without specifying a particular logic, they used operator out to refer to any input/output logic. Operation $out(R) \oplus (a, x)$ indicates the expansion of a rule based-system R by a new rule, operation $out(R) \ominus (a, x)$ denotes the contraction of a rule (a, x) from $out(R)$, and operation $out(R) \otimes (a, x)$ indicates the revision of $out(R)$ by new rule (a, x) .

Like AGM expansion, the definition of rule expansion is straightforward. The new rule that is enforced does not cause any conflict with the existing legal code. Hence, rule (a, x) is added to $out(R)$ together with all the rules that can be derived from the union of $deriv(R)$ and (a, x) : $out(R) \oplus (a, x) = out(R \cup \{(a, x)\})$.

Definition 24. [16] *Let out be an input/output logic. A rule contraction operator \ominus satisfies the following postulates:*

R-1: $out(R) \ominus (a, x)$ is closed under out (closure or type)

R-2: $out(R) \ominus (a, x) \subseteq out(R)$ (inclusion or contraction)

R-3: If $(a, x) \notin out(R)$, then $out(R) = out(R) \ominus (a, x)$ (vacuity or min. action)

R-4: If $(a, x) \notin out(\emptyset)$, then $(a, x) \notin out(R) \ominus (a, x)$ (success)

R-5: If $(a, x) \in out(R)$, then $out(R) \subseteq (out(R) \ominus (a, x)) \oplus (a, x)$ (recovery)

R-6: *If $out(\{(a, x)\}) = out(\{(b, y)\})$, then $out(R) \ominus (a, x) = out(R) \ominus (b, y)$ (extensionality)*

As we have seen in definition 6, the last two AGM postulates ((- 7)-(- 8)) are optional and refer to conjunctions. Since conjunctions are not defined for rules, we restrict ourselves to the basic postulates.

A few words are due about the success postulate. The *success* postulate for rule contraction says that if $x \notin out(\emptyset, a)$, then $x \notin out(R \ominus (a, x), a)$. There are several ways in which a set of rules can be contracted. The purpose of the postulates is to distinguish admissible solutions from inadmissible ones. However, unlike in AGM theory revision, the question here concerns not only what and how much to contract, but also *which inputs* to contract. Boella *et al.* [16] show with the aid of an example that sometimes, in order to obtain a rule-based system that satisfies the success postulate, one needs to *add* some rules.

Another issue is the characterisation of the minimal rule contraction operators. We have seen that in AGM, one interpretation of the postulates is to impose the economical principle. That is, when performing a rule contraction operator, we want to keep as much as possible. However, a syntactic characterisation of minimal rule contraction encounters some problems. In AGM, thanks to the closure postulate (i.e. belief sets are closed under consequence), if $y \notin (K - x)$, then we also have that $x \wedge y \notin (K - x)$. Likewise, if $(a, x) \notin out(R) \ominus (a, x)$, then also $(a, x \wedge y) \notin out(R) \ominus (a, x)$. However, this is not the only consequence of the success postulate for rule contraction. For example, for all six input/output logics considered here, if $(a, x) \notin out(R) \ominus (a, x)$, then also $(a \vee b, x) \notin out(R) \ominus (a, x)$.

Other logical relations depend on the input/output logic used. For example, for basic output out_2 , if $(a, x) \notin out(R) \ominus (a, x)$, then we have either $(a \wedge b, x) \notin out(R) \ominus (a, x)$ or $(a \wedge \neg b, x) \notin out(R) \ominus (a, x)$. In other words, if $(a, x) \notin out(R) \ominus (a, x)$ and $(a \wedge b, x) \in out(R) \ominus (a, x)$, then $(a \wedge \neg b, x) \notin out(R) \ominus (a, x)$. These relations do not hold for simple-minded output out_1 . Likewise, a similar property based on the inverse of CTA holds for reusable output out_3 .

The recovery postulate states that contracting a rule-based system by (a, x) and then expanding by the same (a, x) should leave $out(R)$ unchanged. We will see that such a postulate turns out to be problematic for rule contraction.

Boella *et al.* [16] show that the five postulates considered so far are consistent only for some input/output logics, and not for others. In particular, if we adopt output out_1 or out_3 , then there is no single

Proposition 25. [16]

(R-1) to (R-5) cannot hold together for out_1 or out_3 , but they can hold together for out_2 .

We now turn to the postulates for rule revision.

3.4 Revision of Normative Systems

As in rule contraction, we consider only the first six AGM revision postulates and the rule revision postulates.

Definition 26. [16] Let out be an input/output logic, and $deriv(R)$ a set of rules closed under out . A rule revision operator \otimes satisfies the following postulates:

R \otimes 1: $out(R) \otimes (a, x)$ is closed under out (closure or type)

R \otimes 2: $(a, x) \in (out(R) \otimes (a, x))$ (success)

R \otimes 3: $out(R) \otimes (a, x) \subseteq out(R) \oplus (a, x)$ (inclusion)

R \otimes 4: If $(a, \neg x) \notin out(R \cup (a, x))$ then $out(R) \oplus (a, x) = out(R) \otimes (a, x)$ (vacuity)

R \otimes 5: $(a, \neg x) \in out(R) \otimes (a, x)$ iff $(a, \neg x) \in out(\emptyset)$ (triviality)

R \otimes 6: If $out(\{(a, x)\}) = out(\{(b, y)\})$, then $out(R) \otimes (a, x) = out(R) \otimes (b, y)$ (extensionality)

As seen in section 2, the Levi Identity defines revision $K * A$ as a sequence of a contraction and a expansion. We have seen the correctness of such a definition in observations 10 and 12.

It is worth noting that the controversial recovery postulate (– 5) was not used in observation 12. Boella *et al.* [16] show that the same result can be proven for rule change.

Theorem 27. [16] Given a rule contraction operator, we can define a rule revision operator via the Levi Identity:

$$out(R) \otimes (a, x) = (out(R) \ominus (a, \neg x)) \oplus (a, x).$$

When operator \ominus satisfies rules (R-1) to (R-4) and (R-6), then operator \otimes satisfies rules (R*1) -(R*6).

Not only can belief revision be defined in terms of belief contraction operators, belief contractions can also be defined in terms of belief revisions using the Harper and Levi Identities introduced in section 2 .

However, as recalled in proposition 25, for out_1 and out_3 the revision postulates are consistent and the contraction postulates are not. Thus, a result like observation 10 for normative change does not hold.

We recall from section 2 that the Levi and Harper Identities have been shown to be interchangeable in AGM theory. So, even though there is no theorem corresponding to observation 10 in the general case, one may want to check whether $out(R) \circledast (a, x) = (out(R) \cap out(R) \circledast (a, x)) \oplus (a, x)$ is a consequence of the basic postulates for rule revisions, and whether $out(R) \ominus (a, x) = out(R) \cap ((out(R) \ominus (a, x)) \oplus (a, \neg x))$ can be proven from the basic set of postulates for rule contractions (including the recovery postulate). Boella *et al.* [16] show that the answer to the first question is positive:

Proposition 28. [16] $out(R) \circledast (a, x) = (out(R) \cap out(R) \circledast (a, x)) \oplus (a, x)$.

However, $out(R) \ominus (a, x) = out(R) \cap ((out(R) \ominus (a, x)) \oplus (a, \neg x))$ does not hold in general, i.e. it cannot hold for output out_1 or out_3 .

3.5 Contraction of Normative Bases

Models of belief contraction and revision are built in order to satisfy the demand for minimal change to keep a theory consistent. As we have seen in section 2.1 above, there are two basic strategies for reaching this goal with the syntactic approach. The first consists in selecting the resulting contraction or revision among maximal consistent subsets of the original. The second consists in making an “incision” in the minimal subsets of the theory or base that derived the sentence to be deleted or revised. We shall now follow the second strategy, calling those minimal subsets “arguments”, which are here the base of normative entailments from the set of rules. The construction proceeds basically by making minimal withdrawals from those arguments:

Definition 29. (Argument) $X \subseteq L \times L$ is an argument for (a, x) based on a normative set N if and only if:

- (i) $X \subseteq N$;
- (ii) $(a, x) \in out(X)$;
- (iii) if $X' \subset X$, then $(a, x) \notin out(X')$.

$Args_N(a, x)$ is the set of arguments for (a, x) based on N .

Definition 30. An incision σ is a choice-like function on $\text{Args}_N(a, x)$ to $\wp(L \times L)$ such that:

- (i) $\sigma(\text{Args}_N(a, x)) \subseteq \bigcup \text{Args}_N(a, x)$;
- (ii) $\sigma(\text{Args}_N(a, x)) \cap X \neq \emptyset$, for all $X \in \text{Args}_N(a, x)$.

Definition 31. Let N be a normative set and (a, x) a conditional norm. Then, the contraction of N by (a, x) is defined as:

$$N -_{\sigma} (a, x) = N \setminus \sigma(\text{Args}_N(a, x)).$$

The contraction of a normative set N by a conditional rule (a, x) may also be defined by postulates on a contraction function, as follows.

Definition 32. The contraction of a normative set N by a conditional rule (a, x) is a function $N - : L \times L \rightarrow \wp(L \times L)$ satisfying the following postulates:

- N-1:** if $(a, x) \notin \text{out}(\emptyset)$, then $(a, x) \notin \text{out}(N - (a, x))$ (success)
- N-2:** $N - (a, x) \subseteq N$ (inclusion)
- N-3:** if $(b, y) \in N \setminus N - (a, x)$, then there is $N' \subset N$ such that $(a, x) \notin \text{out}(N')$, but $(a, x) \in \text{out}(N' \cup \{(b, y)\})$ (core-retainment)
- N-4:** if for all $N' \subseteq N$, $(a, x) \in \text{out}(N')$, if and only if $(b, y) \in \text{out}(N')$, then $N - (a, x) = N - (b, y)$ (uniformity)

The representation theorem below is easily adapted from Hansson’s representation theorem for base contraction (observation 15):

Theorem 33. $N -_{\sigma} (a, x) = N - (a, x)$.

3.6 Refinement of Normative Bases

As we have noticed above for output operators stronger than basic output out_2 , the following property holds: if $(a, x) \notin \text{out}(R) \ominus (a, x)$, then either $(a \wedge b, x) \in \text{out}(R) \ominus (a, x)$ or $(a \wedge \neg b, x) \in \text{out}(R) \ominus (a, x)$. Hence, in every contraction of a conditional obligation (a, x) from a closed normative set R , based on an underlying logic at least as strong as basic output, the resulting contracted set $\text{out}(R) \ominus (a, x)$ will include a “weakened” version of the conditional, that is, either $(a \wedge b, x)$ or $(a \wedge \neg b, x)$. It is possible to specify in the selection function which weakened version shall remain. This was the basic intuition underlying the operator called *refinement* proposed by

Maranhão [47], which was aimed at modelling the introduction of exceptions to rules by legal interpretation. For instance, given a normative system that delivers absolute prohibition of abortion, $(\top, \neg\text{abort}) \in O$, a defence of abortion in the case of an anencephalic foetus would not be a proposal for permitting abortion in any context. Hence, the contraction of $(\top, \neg\text{abort})$ from that system should make reference to that specific exception, which means that in the absence of anencephaly, abortion should remain forbidden in that normative system, in the name of minimal change. That is, $(\neg\text{anenceph}, \neg\text{abort})$ should still be derivable from normative system O , while the prohibition should cease to hold in the exceptional case, that is $(\text{anenceph}, \neg\text{abort}) \notin \text{out}(O)$.

By specifying the exception in the selection function, this result follows from the principle of minimality if the normative set is closed and the logic is at least as strong as a basic output. However, for normative bases (not closed sets), deleting $(\text{anenceph}, \neg\text{abort})$ from the set of consequences of O would be tantamount to excluding $(\top, \neg\text{abort})$ from normative set O , and therefore $(\neg\text{anenceph}, \neg\text{abort})$ would not be derived anymore.

But it is possible to define a refinement as a particular case of a *conservative contraction* [49]. That is, it expands the normative set with rules that are entailed by the rule to be contracted, and which include the exceptional factor and its negation in the antecedent.

Definition 34. (*Refinement*) Let $f \in L$, N be a normative system and let $(a, x) \in \text{out}(N)$, where out is at least as strong as a basic output. Then, the refinement of N and (a, x) by factor f is $N \otimes^f (a, x) = N^* -_{\theta_{N^*}} (a, x)$ where $N^* = N \cup \{(f \wedge a, x), (\neg f \wedge a, x)\}$ and $(\neg f \wedge b, y) \notin \theta(\text{Args}_{N^*}(a, x))$. We call factor f an *exception* to (a, x) in the resulting refined normative system.

Proposition 35. *The refinement operator satisfies the following success properties:*

- $(a, x) \notin N \otimes^f (a, x)$;
- $(a, x), (f \wedge a, x) \notin N \otimes^f (a, x)$;
- $(\neg f \wedge a, x) \in N \otimes^f (a, x)$.

3.7 Contraction of Combined Normative Bases

As we have seen in section 3.2, the combination of a constitutive set of rules and regulative sets of permissions and obligations may give rise to genuine incoherencies, that is, the delivery of incompatible rulings, even though the sets of obligations and permissions are consistent. This happens when a given input activates definitions in

the constitutive set that triggers logically independent rules with conflicting outputs. As we have suggested, restoring coherence would involve deciding between several alternatives that may change the set of constitutive rules, or the set of regulative rules, or both. In this section, we are going to introduce a formal framework for the operation of contracting normative systems that combine sets of constitutive rules (which we shall call a *constitutive set*) and regulative rules (which we shall call a *regulative set*).

For $A \subseteq L$, output operators out_i and out_j , constitutive set C and regulative set R , we shall use the following conventions:

- (i) $out_i(C, R, A)$ if $i = j$;
- (ii) $out_{ij}(C, R, a)$ denoting $out_{ij}(C, R, \{a\})$;
- (iii) $(a, x) \in out_{ij}(C, R)$ if $x \in out_{ij}(C, R, a)$.

We call the pair of normative sets (C, R) the combination of C and R or the combination (C, R) .

Below, we build and characterise operators to perform the three kinds of changes in normative systems that combine constitutive and regulative rules. The first operator, called *constitutive contraction*, contracts only the constitutive set. The second operator, called *regulative contraction*, contracts the regulative set. The *combined contraction* operator may contract both in order to delete a norm from the combination of the constitutive and regulative sets.

Definition 36. (*Constitutive contraction*) *The constitutive contraction of a combination (C, R) by a conditional norm (a, x) is a function $C -_R : L \times L \rightarrow \wp(L \times L)$ satisfying the following postulates:*

- C-1:** *if $(a, x) \notin out_i(\emptyset, R)$, then $(a, x) \notin out_i(C -_R(a, x), R)$ (success)*
- C-2:** $C -_R(a, x) \subseteq C$ (inclusion)
- C-3:** *if $(b, y) \in C \setminus C -_R(a, x)$, then there is $C' \subset C$ such that $(a, x) \notin out_i(C', R)$, but $(a, x) \in out_i(C' \cup \{(b, y)\}, R)$ (core-retainment)*
- C-4:** *if for all $C' \subseteq C$ it is the case that $(a, x) \in out_i(C', R)$ if and only if $(b, y) \in out_i(C', R)$, then $C -_R(b, y) = C -_R(a, x)$ (uniformity)*

Definition 37. (*Regulative contraction*) *The regulative contraction of a combination C, R by a conditional norm (a, x) is a function $R -_C : L \times L \rightarrow \wp(L \times L)$ satisfying the following postulates:*

- R-1:** if $(a, x) \notin out_i(C, \emptyset)$, then $(a, x) \notin out_i(C, R -_C(a, x))$ (success)
- R-2:** $R -_C(a, x) \subseteq R$ (inclusion)
- R-3:** if $(b, y) \in R \setminus R -_C(a, x)$, then there is an $R' \subset R$ such that $(a, x) \notin out_i(C, R')$, but $(a, x) \in out_i(C, R' \cup \{(b, y)\})$ (core-retainment)
- R-4:** if for all $R' \subseteq R$, $(a, x) \in (C, R')$ if and only if $(b, y) \in out_i(C, R')$, then $R -_C(a, x) = R -_C(b, y)$ (uniformity)

We use the following conventions for the definition of the combined contraction of normative sets:

- (i) if $(C, R) - (a, x) = (C^-, R^-)$, then $(C, R) \setminus (C, R) - (a, x) = (C \setminus C^-, R \setminus R^-)$;
 (ii) $\bigcup(C, R) = \bigcup\{C, R\}$.

Definition 38. (Combined contraction) The combined contraction of the combination (C, R) by a conditional norm (a, x) is a function $(C, R) - : L \times L \rightarrow \wp(L \times L) \times \wp(L \times L)$ satisfying the following postulates:

- C/R-1:** if $(a, x) \notin out_i(\emptyset)$, then $(a, x) \notin out_i((C, R) - (a, x))$ (success)
- C/R-2:** if $(C, R) - (a, x) = (C^-, R^-)$, then $C^- \subseteq C$ and $R^- \subseteq R$ (inclusion)
- C/R-3:** if $(b, y) \in \bigcup(C, R) \setminus (C, R) - (a, x)$, then there is a $C' \subseteq C$ and $R' \subseteq R$ such that $(a, x) \notin out(C', R')$, but $(a, x) \in out_i(C' \cup \{(b, y)\}, R')$ or $(a, x) \in out_i(C', R' \cup \{(b, y)\})$ (core-retainment)
- C/R-4:** if for all $C' \subseteq C$ and $R' \subseteq R$, it is the case that $(a, x) \in out_i(C', R')$ if and only if $(b, y) \in out_i(C', R')$, then $(C, R) - (a, x) = (C, R) - (b, y)$ (uniformity)

Now we will define a general construction for kernel contraction of combined normative sets, from which we may specify constitutive, regulative and combined contraction operators.

Definition 39. (Combined argument) A combination (X, Y) is a combined argument for (a, x) based on the combination (C, R) of a constitutive set C and a regulative set R if and only if:

- (i) $X \subseteq C$;
 (ii) $Y \subseteq R$;
 (iii) $(a, x) \in out_i(X, Y)$;
 (iv) if $X' \subset X$, then $(a, x) \notin out_i(X', Y)$;
 (v) if $Y' \subset Y$, then $(a, x) \notin out_i(X, Y')$.

We denote by $Args_{(C,R)}(a, x)$ the set of combined arguments for (a, x) based on (C, R) . Now we will define the incision function for choosing rules from the minimal arguments delivering the rule to be excluded.

Definition 40. *An incision is a choice-like function on $Args_{(C,R)}(a, x)$ to $\wp(L \times L)$ such that:*

- (i) if $Args_{(C,R)}(a, x) = \{(X_i, Y_i) : i \in I\}$,
then $\sigma(Args_{(C,R)}(a, x)) \subseteq \bigcup_{i \in I} (X_i \cup Y_i)$;
- (ii) $\sigma(Args_{(C,R)}(a, x)) \cap (X_i \cup Y_i) \neq \emptyset$ for every $(X_i, Y_i) \in Args_{(C,R)}(a, x)$.

The general definition encompasses incisions that choose rules from both normative sets at the same time, incisions that choose only regulative rules, and incisions that choose only constitutive rules. The definitions above restrict the incision functions to choosing only constitutive rules or only regulative rules.

Definition 41. *An incision on $Args_{(C,R)}(a, x)$ is constitutive if and only if $\sigma(Args_{(C,R)}(a, x)) \cap R = \emptyset$.*

Definition 42. *An incision on $Args_{(C,R)}(a, x)$ is regulative if and only if $\sigma(Args_{(C,R)}(a, x)) \cap C = \emptyset$.*

Now we will use a general definition for contraction based on the incision function. Of course, if we use a constitutive incision, the result will be a constitutive contraction. Similarly, if we use a regulative incision, the result will be a regulative contraction.

Definition 43. *(Contraction) Let (C, R) be a combination of normative sets and (a, x) a conditional norm. Then, the contraction of (C, R) by (a, x) based on incision σ is defined as $(C, R) -_{\sigma} (a, x) = (C^-, R^-)$ where $C^- = C \setminus \sigma(Args_{(C,R)}(a, x))$ and $R^- = R \setminus \sigma(Args_{(C,R)}(a, x))$.*

The theorems below show that the postulates for constitutive, regulative and general contraction characterise the respective constructions.

Theorem 44. *[52] A contraction of (C, R) by (a, x) based on a constitutive incision σ is a constitutive contraction, that is, $(C, R) -_{\sigma} (a, x) = (C -_R(a, x), R)$. Moreover, given a constitutive contraction, there is a constitutive incision σ such that $(C, R) -_{\sigma} (a, x) = (C -_R(a, x), R)$.*

Theorem 45. *[52] A contraction of (C, R) by (a, x) based on a regulative incision σ is a regulative contraction, that is, $(C, R) -_{\sigma} (a, x) = (C, R -_C(a, x))$. Moreover, given a regulative contraction, there is a regulative incision σ such that $(C, R) -_{\sigma} (a, x) = (C, R -_C(a, x))$.*

Theorem 46. [52]

$$(C, R) -_{\sigma} (a, x) = (C, R) - (a, x).$$

The contraction operators discussed here do not involve constraints on the choice of incision function that will determine the result of the contraction operation. Therefore, there is no preference for a regulative contraction over a constitutive or combined contraction.

This feature may be illustrated by example 4, which was formalised in section 3.2. In that case, a contraction to avoid $sms \wedge sord \wedge \neg iord$ -incoherent would have the following alternatives for the incisions: $(C, O) -_{\sigma} (sms \wedge sord \wedge \neg iord, \neg acc)$ or $(C, P) -_{\sigma} (sms \wedge sord \wedge \neg iord, acc)$, each of which is determined by any of the following unitary incision functions: $\sigma_1 = \{(sms, dat)\}$, or $\sigma_2 = \{(sms, com)\}$, or $\sigma_3 = \{(data, prop)\}$, or $\sigma_4 = \{(prop \wedge sord, acc)\}$, or $\sigma_5 = \{(com \wedge \neg iord, \neg acc)\}$.

The controversy within the Brazilian Superior Court of Justice discussed in section 1.2 involved two of these alternative contractions. The first decision was a constitutive contraction based on σ_1 , where the court contended that message exchanges are communications in flux, which demanded a specific order to intercept the conversation.

In turn, the second decision by the Brazilian court was a conservative contraction based on σ_2 , contending that message exchanges should not be considered as ongoing communication. The same alternative contraction was chosen by the German court. The underlying reason for these choices was the weight given to the constitutional value of freedom of communication, which is demoted by such access to the content of an individual's mobile phone. The demotion of freedom of communication was considered stronger than the demotion of property rights. Hence, the association of "text messaging" with "stored data" and, therefore, with "property" (instead of its association with "personal communication") coheres with an underlying valuation where property rights are outweighed by public safety concerns. The German decision also involved a concern about the constitutional right of informational autonomy as the core of data protection. According to the court's argumentation, this right was not violated because the data subject could have destroyed the data in her possession.

Notice that both courts decided not to revise the regulative rules, only stipulate the conceptual qualification of text messaging. The contraction of the regulative set would be inadequate. The first alternative contraction, σ_4 , would lead to the absence of an explicit authorisation to search property items, while the other alternative contraction, σ_5 would exclude the prohibition to intercept communications. Nevertheless, the court could have considered less intrusive interventions on the set

of regulative rules by, for instance, treating the case of text messaging as an exception to search orders on data. That is, in order to reach a coherent normative system in that context (to avoid $sms \wedge sord \wedge \neg iord$ -incoherence), the court could have refined the set of obligations, which in the model would be represented by a refinement operator ensuring that $(\neg sms \wedge prop \wedge sorder, acc) \in P \otimes^{sms} (prop \wedge sorder, acc)$.

The resulting contraction would then be either constitutive or regulative. However, there can be genuine combined contractions on sets of constitutive and regulative rules. Consider, for instance, a variation on example 4, where an order to investigate an individual (*order*) would encompass both a search & seizure procedure and the interception of any communication. We would have the following sets in the normative system:

$$C = \{(sms, dat), (data, prop), (sms, com)\}$$

$$P = \{(com \wedge order, acc), (prop \wedge order, acc)\}$$

According to that normative system, police officers are authorised to access the content of the message exchange stored on the cell phone with a general order authorising the investigation of an individual. Now suppose that the legislator derogates from the positive permission to access the content of text messages stored on a mobile phone, or that legal interpretation (judicial or doctrinal) considers such a permission to be unconstitutional for violating the fundamental right to privacy. In that case, a contraction $(C, P) - (sms \wedge order, acc)$ involves choosing from the following incisions:

$$\sigma_1 = \{(sms, dat), (com, acc)\}$$

$$\sigma_2 = \{(dat, prop), (com, acc)\}$$

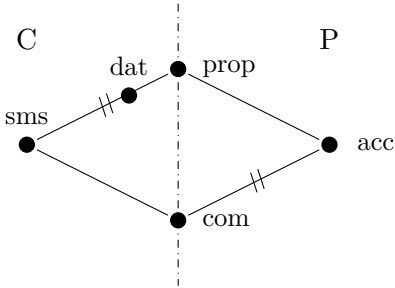
$$\sigma_3 = \{(sms, com), (prop, acc)\}$$

$$\sigma_4 = \{(sms, com), (sms, dat)\}$$

$$\sigma_5 = \{(sms, com), (dat, prop)\}$$

$$\sigma_6 = \{(prop, acc), (com, acc)\}$$

The contractions based on σ_{1-3} are combined contractions, while those based on σ_{4-5} are constitutive contractions. The contraction based on σ_6 is the only alternative based on regulative contraction. The figure below illustrates incision σ_1 , where each dash linking two nodes is a pair, and each node is proposition:



The contraction based on σ_1 is indeed the most reasonable choice. A regulative contraction is clearly undesirable, since it would make any search unauthorised, resulting in a normative system that completely disregards the value of public safety and containing useless definitions. On the other hand, a constitutive contraction based on σ_4 would not address the crucial question in this case, which is how to legally qualify text messaging. In turn, the constitutive contraction σ_5 would make it impossible to search for any document on the premises, in spite of defining that text messaging counts as data. The combined contraction σ_2 would be similar to σ_1 , with the effect of favouring freedom of communication over public safety. However, it would also have the undesirable effect of hindering access to any data in a search procedure. For a similar reason, σ_3 would be inadequate with regard to the intuition that the protection of property rights has less weight than the protection of freedom of communication when balancing public safety concerns.

4 Challenges and Open Problems with the AGM Approach

In this final section, we will discuss some open problems and relevant questions that are the object of mainstream research on normative change with the AGM approach.

As we have seen in section 2.2, one of the main challenges and criticisms of the AGM approach is the potential indeterminacy of the result of a contraction, revision or refinement of the normative system, which depends on choices about the proper selection or incision functions to determine the result. This feature is sometimes seen as a disadvantage compared to the syntactic approach, where the syntax of a particular rule is the object of change.

Actually, as we have argued in section 1.1, what we have called the “indeterminacy problem” is not really a defect of the representational model, but is a real

feature of legal reasoning about normative change that should be captured by the model itself. As a representation of the activity of legal interpretation, it is particularly interesting to show what are the alternative interpretations for different acts of the derogation, making it clear that a particular interpretation involves choices.

Although there may be some alternative interpretations that are clearly inadequate and would be immediately rejected by a jurist, it is important to investigate the criteria for rejection and represent them in the model. It is also a fact that there may be a doctrinal or judicial controversy concerning the defensible results of a normative change, as illustrated in example 4, and we believe that the model should be able to express these different available choices as an adequate representation of legal reasoning. So we see the indeterminacy reflected in the model as an advantage of the AGM approach.

However, there is also an *onus* on this model to provide criteria that would reflect the consensual choices (in the sense of consensus on action, not consensus on explicit convention) reached by legal practitioners and jurists on normative change. Hence, one of the main challenges to research on normative change is to find and model criteria for determining rational choices from alternative normative systems resulting from change operations.

When discussing the examples formalised in section 3.7, we provided some reasons for preferring certain incisions over others. The arguments used there to justify the choice of a particular incision were all domain-specific. Nevertheless, the discussion provided at least two important clues for developing more abstract constraints.

The first clue is related to Makinson and van der Torre's discussion on constraints for I/O-logics [43] suggesting a distinction between rule maximisation (*maxrule*: maximising the preservation of rules in order to satisfy a constraint) and output maximisation (*maxout*: maximising the preservation of outputs in order to satisfy a constraint). The Mobius Strip example is a radical case and may be seen as a contraction. Consider $N = \{(\top, a), (a, b), (b, \neg a)\}$. The contraction $N - (\top, \perp)$ has two possible outcomes: $N_1 = \{(\top, a)\}$ or $N_2 = \{(a, b), (b, \neg a)\}$. While N_1 satisfies *maxout* and fails *maxrule*, N_2 satisfies *maxrule* and fails *maxout*.

Indeed, constitutive contractions tend to favour *maxrule* and sacrifice *maxout*, since intermediary concepts may be connected to different rules. As we have indicated in section 1.1, the network effects problem regarding normative change alerts us that suppressing relevant connections between normative concepts may render regulative rules inapplicable, while deleting regulative rules may change our understanding of some normative concepts. The construction of the contraction operators for combined normative sets in this article was based on rule maximisation, but future investigations should try to find reasonable constraints to temper the demand for *maxrule* with the demand for *maxout*.

The second clue is the role of values that drive the choices among possible outcomes of a change function. The positively enacted rules (constitutive and regulative) on which the legal order are built are the outcomes of (legislative or judicial) deliberations on relevant societal values (moral considerations, political goals, fundamental rights). Those societal values inform the interpretation of authoritative decisions in the application of the rules of the normative system when evaluating the legality of actions in particular contexts. Such values may be considered as external to the normative system or as internal to it in the form of constitutional rights and principles. Thus, if one conceives of legal interpretation as a dynamic of normative change, as suggested in section 1.2, then enriching the model with reasoning about balancing values would provide relevant criteria for choosing between the resulting contracted, revised or refined normative systems, a line of research recently pursued by Maranhão [50] and Maranhão and Sartor [53].

If one takes seriously the representation of legal interpretation as normative change, and succeeds in modelling relevant criteria for choosing among possible systems resulting from contractions, revisions and refinements, then argumentation frameworks could be developed to model argumentation by legal doctrine to determine the best interpretation. That is, there could be a model of argumentation about the results of normative change. Such an argumentation process would put forward defeasible arguments about competing goals of legal interpretation (consistency, coherence with underlying political morality, completeness, precision, adherence to positively enacted rules and natural language, etc.).

The incorporation of tools to represent reasoning about values in the model of normative change will inevitably lead to the need to adapt the change functions to non-monotonic logic, including input/output logics where its rules are default (see [56]). There is a fairly dominant trend in legal theory [9] and in the literature of artificial intelligence & law (see [12], [61] and [69]) of considering reasoning about values as defeasible, where consideration of additional values in a particular context may defeat reasons for particular actions in a framework of an overall appreciation of those relevant values. Hence, as already mentioned in section 2.2, the AGM methodology should be adapted to systems with underlying logics that are not monotonic, as pursued recently by Zhuang *et al.* [79], Casini and Meyer [20], and Casini *et al.* [19]. Since the addition of new values or considerations related to values may defeat some implications or reasons for action, with the AGM approach such systems will reflect an aspect of the syntactic approach where a “contraction” is obtained by adding rules to the normative set [50]. As argued in sections 1.2 and 1.4, the representation of legal interpretation should involve values, and that aspect may point to incorporating methods of revision provided by what we have called the preferential approach.

Another important observation concerning applications of the formal models of normative change, emphasised in sections 1.2 and 1.4, is how to adequately represent the two dimensions of normative change: the dimension of validity, which we believe is better reflected by the AGM approach, and the dimension of efficacy, which seems to be better captured by the syntactic approach. Integrating both perspectives would also demand formal comparisons between these approaches. Where the AGM approaches focus on changes in the normative system, it is pertinent to ask whether and how the resulting system can be captured by syntactic modifications of the rules and how alternative interpretations can be represented. Where the syntactic approaches focus on the syntactical representation of the time span of the efficacy of rules and how to block or enable their effects, it is pertinent to ask whether the enabled rules in a given time span can be represented by a temporal dynamic for subsystems of the whole system of valid rules (containing the rules that are enabled at a given period). Efforts to enrich the syntactic representation of rules within the AGM approach with, for instance, time labels [74], are also important for modelling reasoning that closely reflects real-life examples of the complex interaction between the period of a rule's efficacy and the time span of its validity in the legal system.

There are also conceptual and formal results to be pursued by researchers working on the AGM approach. For instance, there are still no formal characterisations of revision and refinement for changing combined normative sets. It is also relevant to explore constructions of revision from contraction and vice versa for some input/output logics where the Harper and Levi Identities would not hold (see section 3.4). A general theory of revision functions on different sorts of architectures of input/output logics (combinations of normative sets within the input/output logics framework) would also be a relevant theoretical achievement to ground future research of applications that explore particular architectures [17, 53] for more complex architectures).

The constructions discussed in this article were based on original input/output logics (simple-minded, basic, reusable and basic reusable) introduced by Makinson and van der Torre [42]. It would be interesting to apply the AGM approach to input/output logics with constraints [43] and other variants [58, 59].

Acknowledgments

Leendert van der Torre acknowledges financial support from the Fonds National de la Recherche Luxembourg (INTER/Mobility/19/13995684/DLAI/van der Torre). Giovanni Casini has been supported by TAILOR, a project funded by EU Horizon 2020 research and innovation programme under GA No 952215.

References

- [1] Aulis Aarnio. *On Legal Reasoning*. Turun Yliopisto, 1977.
- [2] Carlos E Alchourrón. Logic of Norms and Logic of Normative Propositions. *Logique et analyse*, 12:242–268, 1969.
- [3] Carlos E. Alchourrón. Detachment and Defeasibility in Deontic Logic. *Studia Logica*, 57:5–18, 1996.
- [4] Carlos E Alchourrón and Eugenio Bulygin. *Normative Systems*. Springer, 1971.
- [5] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.
- [6] Carlos E. Alchourrón and David Makinson. *Hierarchies of Regulations and their Logic*, pages 125–148. Springer Netherlands, 1981.
- [7] Carlos E. Alchourrón and David Makinson. On the logic of theory change: Contraction functions and their associated revision functions. *Theoria*, 48:14–37, 1982.
- [8] Robert Alexy. On the structure of legal principles. *Ratio Juris*, 13:294–304, 2000.
- [9] Robert Alexy. Constitutional Rights, Balancing, and Rationality. *Ratio Juris*, 16:131–140, 2003.
- [10] Amalia Amaya. *The Tapestry of Reason: an inquiry into the nature of coherence and its role in legal argument*. Hart Publishing, 2015.
- [11] Aharon Barak. *Purposive Interpretation in Law*. Princeton University Press, 2005.
- [12] Trevor J. M. Bench-Capon and Giovanni Sartor. A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*, 150:97–142, 2003.
- [13] Donald H. Berman and Carole D. Hafner. Representing teleological structure in case-based reasoning: The missing link. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law (ICAIL)*, pages 5–9. ACM, 1993.
- [14] Norberto Bobbio. Le bon législateur. *Logique & Analyse*, 14(53-54):243–249, 1971.
- [15] Guido Boella, Guido Governatori, Antonino Rotolo, and Leendert van der Torre. Lex minus dixit quam voluit, lex magis dixit quam voluit: A formal study on legal compliance and interpretation. In *AICOL-I/IVR-XXIV'09 Proceedings of the 2009 international conference on AI approaches to the complexity of legal systems: complex systems, the semantic web, ontologies, argumentation, and dialogue*, pages 162–183. Springer, 2010.
- [16] Guido Boella, Gabriella Pigozzi, and Leendert van der Torre. AGM contraction and revision of rules. *Journal of Logic, Language and Information*, 25(3-4):273–297, 2016.
- [17] Guido Boella and Leendert van der Torre. A Logical Architecture of a Normative System. In *Deontic Logic and Artificial Normative Systems - DEON 2006*, pages 24–35. Springer, 2006.
- [18] Guido Boella and Leendert van der Torre. A logical architecture of a normative system. In *Deontic Logic and Artificial Normative Systems - DEON 2006*, pages 24–35. Springer, 2006.

- [19] Giovanni Casini, Eduardo Fermé, Thomas Meyer, and Ivan Varzinczak. A semantic perspective on belief change in a preferential non-monotonic framework. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018.*, pages 220–229, 2018.
- [20] Giovanni Casini and Thomas A. Meyer. Belief change in a preferential non-monotonic framework. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 929–935, 2017.
- [21] James P. Delgrande and Pavlos Peppas. Belief revision in Horn theories. *Art. Int.*, 218:1 – 22, 2015.
- [22] Julie Dickson. Interpretation and Coherence in Legal Reasoning. In Edward N Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2016.
- [23] Ronald M Dworkin. *Law’s Empire*. Kermode, 1986.
- [24] Eduardo L. Fermé. On the logic of theory change: Contraction without recovery. *Journal of Logic, Language and Information*, 7(2):127–137, Apr 1998.
- [25] Eduardo L. Fermé and Sven O. Hansson. *Belief Change - Introduction and Overview*. Springer Briefs in Intelligent Systems. Springer, 2018.
- [26] Lon L Fuller. Positivism and Fidelity to Law - A Reply to Professor Hart. *Harvard Law Review*, 71:630–672, 1958.
- [27] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, 1988.
- [28] Peter Gärdenfors and Hans Rott. Belief revision. In C.J. Hogger D.M. Gabbay and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume IV: Epistemic and Temporal Reasoning*, pages 35–132. Oxford University Press, 1995.
- [29] Guido Governatori, Francesco Olivieri, Matteo Cristani, and Simone Scannapieco. Revision of defeasible preferences. *Int. J. Approx. Reason.*, 104:205–230, 2019.
- [30] Guido Governatori and Antonino Rotolo. Changing legal systems: legal abrogations and annulments in defeasible logic. *Logic Journal of IGPL*, 18(1):157–194, 2010.
- [31] Guido Governatori, Antonino Rotolo, Francesco Olivieri, and Simone Scannapieco. Legal contractions: a logical analysis. In Enrico Francesconi and Bart Verheij, editors, *ICAIL*, pages 63–72. ACM, 2013.
- [32] Davide Grossi, John-Jules Ch. Meyer, and Frank Dignum. The many faces of counts-as: A formal analysis of constitutive rules. *Journal of Applied Logic*, 6:192–217, 2008.
- [33] Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.
- [34] Sven O. Hansson. Reversing the Levi identity. *Journal of Philosophical Logic*, 22:637–669, 1993.
- [35] Sven O. Hansson. *A Textbook of Belief Dynamics: Theory Change and Database Up-*

- dating*. Kluwer Academic Publishers, 1999.
- [36] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning, KR'91*, pages 387–394, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc.
 - [37] Hirofumi Katsuno and Alberto O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 3(52):263–294, 1991.
 - [38] Hans Kelsen. Derogation. In *Essays in Legal and Moral Philosophy*, pages 261–275. Springer Netherlands, 1973.
 - [39] Sebastien Konieczny and Ramon Pino Perez. Merging information under constraints: A logical framework. *Journal of Logic and Computation*, 12(5):773–808, 10 2002.
 - [40] Isaac Levi. Subjunctives, dispositions and chances. *Synthese*, 34:423–455, 1977.
 - [41] David Makinson. On the status of the postulate of recovery in the logic of theory change. *Journal of Philosophical Logic*, 16(4):383–394, Nov 1987.
 - [42] David Makinson and Leendert van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.
 - [43] David Makinson and Leendert van der Torre. Constraints for input-output logics. *Journal of Philosophical Logic*, 30(2):155–185, 2001.
 - [44] David Makinson and Leendert van der Torre. Permissions from an input-output perspective. *Journal of Philosophical Logic*, 32(4):391–416, 2003.
 - [45] David Makinson and Leendert van der Torre. What is input/output logic? In L'owe Benedikt, Malzkorn Wolfgang, and R'asch Thoralf, editors, *Foundations of the Formal Sciences II: Applications of Mathematical Logic in Philosophy and Linguistics*, pages 163–174. Kluwer, Dordrecht, 2003.
 - [46] Juliano Maranhão. Refinement. A tool to deal with inconsistencies. In *Proc. of the 8th ICAIL*, pages 52–59, 2001.
 - [47] Juliano Maranhao. Some Operators for Refinement of Normative Systems. In *Proceedings of Jurix 2001*, pages 103–115. IOS Press, 2001.
 - [48] Juliano Maranhão. Why was Alchourrón afraid of snakes? *Analisis Filosofico*, XXVI(1):62–92, 2006.
 - [49] Juliano Maranhão. Conservative Contraction. In *The many sides of logic*, pages 465–479. College Publications, 2009.
 - [50] Juliano Maranhão. A logical architecture for dynamic legal interpretation. In *Proceedings of the Eight International Conference on AI and Law ICAIL '17*,, pages 129–38. ACM Press, 2017.
 - [51] Juliano Maranhao. *Positivismo jurídico lógico-incluyente*. Marcial Pons, 2017.
 - [52] Juliano Maranhão and Edelcio G. de Souza. Contraction of combined normative sets. In *Deontic Logic and Normative Systems: 14th International Conference, DEON 2018*, pages 247–261. Springer, 2018.
 - [53] Juliano Maranhão and Giovanni Sartor. Value assessment and revision in legal interpretation. In *Proceedings of the 17th International Conference on Artificial Intelligence and*

- Law, ICAIL 2019*, pages 219–223, New York, New York, USA, jun 2019. Association for Computing Machinery, Inc.
- [54] Andrei Marmor. *Interpretation and Legal theory*. Hart, 2005.
- [55] Pablo E Navarro and Jorge L Rodriguez. *Deontic Logic and Legal Systems*. Cambridge University Press, 2014.
- [56] Xavier Parent. Moral particularism in the light of deontic logic. *Artificial Intelligence and Law*, 19:75–98, 2011.
- [57] Xavier Parent and Leendert van der Torre. *Input/output logics*. College Publications, London, 2013.
- [58] Xavier Parent and Leendert van der Torre. The pragmatic oddity in norm-based deontic logics. In *Proceedings of the International Conference on Artificial Intelligence and Law*, pages 169–178. Association for Computing Machinery, jun 2017.
- [59] Xavier Parent and Leendert van der Torre. - input/output logics with a consistency check. In *Proceedings of the 14th International Conference on Deontic Logic and Normative Systems (DEON2018)*, 2018.
- [60] Aleksander Peczenik. *On Law and Reason*. Kluwer, 1989.
- [61] Henry Prakken, Adam Wyner, Trevor R. Bench-Capon, and Katie Atkinson. A formalisation of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 25:1141–1166, 2015.
- [62] Joseph Raz. Legal Positivism and the Sources of Law. In *The Authority of Law*, pages 37–52. Oxford University Press, 1979.
- [63] Joseph Raz. Authority, law and morality. *The Monist*, 68(3):295–324, jul 1985.
- [64] Joseph Raz. *Authority and Interpretation*. Oxford University Press, 2009.
- [65] Jandson S. Ribeiro, Abhaya Nayak, and Renata Wassermann. Towards belief contraction without compactness. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018*, pages 287–296, 2018.
- [66] Antonino Rotolo. *Identità e somiglianza: saggio sul pensiero analogico nel diritto*. Clueb, 2001.
- [67] Hans Rott. Two dogmas of belief revision. *Journal of Philosophy*, 97(9):503–522, 2000.
- [68] Giovanni Sartor. The Nature of Legal Concepts: Inferential Nodes or Ontological Categories. In Gianmaria Ajani, Ginevra Peruginelli, Giovanni Sartor, and Daniela Tiscornia, editors, *Proceeding of the Conference on “Approaching the Multilanguage Complexity of European Law: Methodologies in Comparison”*. European Press Academic Publishing, 2007.
- [69] Giovanni Sartor. The logic of proportionality: Reasoning with non-numerical magnitudes. *German Law Journal*, pages 1419–57, 2013.
- [70] Giovanni Sartor. Consistency in balancing: from value assessments to factor-based rules. In D. Duarte and S. Sampaio, editors, *Proportionality in Law: An Analytical Perspective*, pages 121–36. Springer, 2018.
- [71] John R Searle. *The Construction of Social Reality*. Free Press, 1995.

- [72] Lawrence B. Solum. The interpretation-construction distinction. *Constitutional Commentary*, 27:95–118, 2010.
- [73] Audun Stolpe. Norm-system revision: theory and application. *Artificial Intelligence and Law*, 18:247–283, 2010.
- [74] Luciano H. Tamargo, Diego C. Martínez, Antonino Rotolo, and Guido Governatori. Temporalised belief revision in the law. In *Frontiers in Artificial Intelligence and Applications*, volume 302, pages 49–58. IOS Press, 2017.
- [75] Luciano H. Tamargo, Diego C. Martínez, Antonino Rotolo, and Guido Governatori. Time, defeasible logic and belief revision: Pathways to legal dynamics. *FLAP*, 8(4):993–1022, 2021.
- [76] Hans Van Ditmarsch and Barteld Kooi. The secret of my success. *Synthese*, 151(2):201–232, Jul 2006.
- [77] Jerzy Wróblewski. Legal Language and Legal Interpretation. *Law and Philosophy*, 4:239–255, 1985.
- [78] Jerzy Wróblewski. *The Judicial Application of Law*. Kluwer, 1992.
- [79] Zhiqiang Zhuang, James P. Delgrande, Abhaya C. Nayak, and Abdul Sattar. Reconsidering AGM-style belief revision in the context of logic programs. In *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*, pages 671–679, 2016.
- [80] Zhiqiang Zhuang, Zhe Wang, Kewen Wang, and Guilin Qi. DL-lite contraction and revision. *J. Artif. Intell. Res.*, 56:329–378, 2016.

MULTI-AGENT ARGUMENTATION AND DIALOGUE

RYUTA ARISAKA

Kyoto University, Kyoto, Japan

ryutaarisaka@gmail.com

JÉRÉMIE DAUPHIN

University of Luxembourg, Esch-sur-Alzette, Luxembourg

jeremie.dauphin@uni.lu

KEN SATOH

National Institute of Informatics, Tokyo, Japan

ksatoh@nii.ac.jp

LEENDERT VAN DER TORRE

University of Luxembourg, Esch-sur-Alzette, Luxembourg

leon.vandertorre@uni.lu

Abstract

This article provides an overview of multi-agent abstract argumentation and dialogue, and its application to formalising legal reasoning. The basis of multi-agent abstract argumentation is input/output argumentation, distinguishing between individual acceptance by agents and collective acceptance by the system. The former may also be seen as a kind of conditional reasoning, and the latter may be seen as the reasoning of an external observer. We extend input/output argumentation in two ways. First, we introduce epistemic trust and agent communication. The former is based on a social network representing epistemic trust, and the latter is based on so-called sub-framework semantics. Second, we introduce dialogue semantics for abstract argumentation by refining agent communication into dialogue steps. A dialogue is a sequence of steps

We thank Massimiliano Giacomin and our two anonymous reviewers for insightful feedback on an earlier version of this article. Leendert van der Torre acknowledges financial support from the Fonds National de la Recherche Luxembourg (INTER/Mobility/19/13995684/ DLAI/van der Torre).

The research carried out for this article was funded from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No. 690974 for the project "MIREL: MIning and REasoning with Legal texts".

from the framework to the extensions, where at each step an agent can commit to accepting some arguments, or commit to hiding or revealing one of his/her rejected arguments. The revealed arguments are then aggregated and an external observer, in our example a judge, can compute which arguments are finally accepted at a global level.

1 Introduction

In his historical overview of formal argumentation, Prakken [42] distinguishes between two kinds of approaches, which he calls argumentation-as-dialogue and argumentation-as-inference. The former is based on protocols and game theory, and the latter is based on non-monotonic logic and graph theory. While in the former approach agents and time play a central role, in the latter approach they are often abstracted away.

In game theory, there is a related distinction between extensive games, which make agents and time explicit, and strategic games, which use the concept of a strategy or conditional plan to abstract time away. The relationship between extensive and strategic games is well understood, in the sense that they are two views of the same phenomenon at different levels of abstraction. This understanding is still missing in the relationship between argumentation-as-dialogue and argumentation-as-inference, despite some work relating these two traditions to the other. For example, Dung [26] shows how his abstract theory can also be applied to reasoning in game theory, and various authors have developed dialogue-based decision procedures for abstract and structured argumentation [19]. We believe that there is a common theory to be developed for argumentation-as-dialogue and argumentation-as-inference, bringing new insights to both. As a first step, we therefore raise the following research question:

Research question. How to introduce agent interaction and dialogue into Dung's abstract argumentation theory?

The starting point for abstract agent argumentation [7], also called triple-A, is the concept of conditional acceptance. In particular, an argument that an agent does not accept can still be put forward as part of the discussion. For example, the agent can explain why (s)he does not accept an argument by presenting counter-arguments to the unaccepted argument, and (s)he may even be willing to accept the argument if convinced by the other agents that his/her counter-arguments are wrong. The following example illustrates how an agent's individual acceptance function is conditional on accepting the arguments of other agents attacking his/her argument, and how this allows us to model one kind of counter-factual argument.

Example 1 (Conditional acceptance). Consider the agent argumentation framework visualised on the left-hand side of Figure 1, whose formal definitions are explained in Section 2. Agent A is an expert in healthcare management considering the argument that a new virus is contained (argument a) and the argument that an additional hospital needs to be built (argument b). Moreover, agent A assumes that if the first argument is accepted, the latter should not be accepted, which is represented by an attack visualised as an arrow from argument a to argument b . Now, consider the multi-agent argumentation framework visualised on the right-hand side of Figure 1. Agent B is an expert in virology who argues that the virus will not be contained (argument c), which attacks argument a of agent A. Since agent A is not an expert in virology, agent A cannot judge whether argument c should be accepted or not.

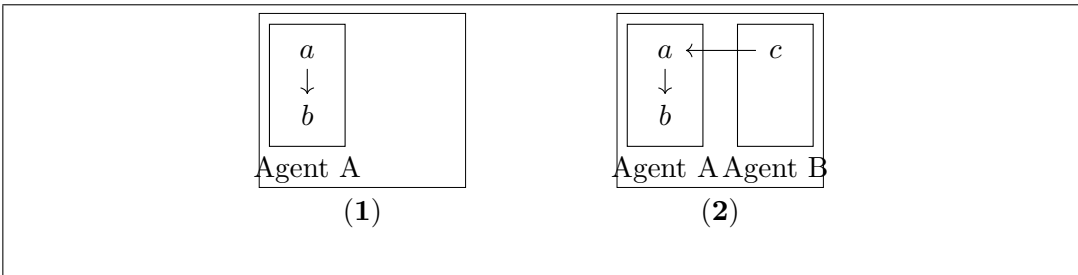


Figure 1: The agent and multi-agent framework of Example 1.

Whether agent A accepts his/her own argument b depends on whether agent B accepts argument c , and on whether agent A trusts agent B on argument c . In particular, if agent B accepts argument c and agent A trusts agent B, then agent A accepts argument b . Otherwise, agent A will accept argument a and reject argument b .

Finally, consider an external observer. Will (s)he accept argument b that an additional hospital must be built? That depends not only on the trust of agent A in agent B, but also whether agent A communicates argument b . In particular, when agent A does not accept argument b , (s)he may decide not to inform agent B or the external observer about the existence of the argument.

To formalise conditional and multi-agent argumentation, abstract agent argumentation uses the theory of input/output argumentation described by Baroni *et al.* [9], also known as multi-sorted argumentation [47]. This theory allows arguments to be assigned to agents, and individual acceptance functions to be associated with these agents. Following the above example, and building on various theories in the literature, abstract agent argumentation extends input/output argumentation in two ways.

First, whether an agent accepts an argument put forward by another agent depends on the trust the agents have in one other, which may be based on their respective reputations [46, 34, 51, 49, 39]. This is often represented by a social network [25, 31, 50], and we follow that tradition in this article. For instance, in Example 1, agent *A* will only reject argument *a* and accept argument *b* if (s)he trusts agent *B* on argument *c*. Agents *A* and *B* may be part of the same coalition cooperating in building a common view of the situation.

Secondly, dialogue is strategic, in the sense that sometimes it is better not to reveal an argument. For example, agent *A* may not like argument *b*, and may thus decide not to reveal this argument to the other agents. This aspect is missing in Dung-style abstract argumentation in the sense that in the dialogue procedures [19], the set of available dialogue actions does not change when arguments are put forward by other agents. Likewise, in most structured argumentation theories, the knowledge bases are assumed to be fixed. Therefore, there is no advantage for an agent in a dialogue game to not put forward an argument. We introduce a new concept: agents can decide whether to hide some of their arguments from the other agents. This concerns, in particular, arguments they do not accept themselves. For example, assume that a scientist knows how increased temperature leads to rising sea levels, but she does not accept this argument herself. She may decide to hide this argument from public debate, because she does not want to give ammunition to her opponents.

From the perspective of an external observer, the interaction is a game between agents. Since the arguments an agent reveals may depend on the arguments revealed by other agents, game-theoretic equilibrium among agents is necessary for the external observer. In a game-theoretic equilibrium, the behaviour of the agents depends on the behaviour of the other agents, in our case for both communicating and accepting arguments. Moreover, even if the agents do not accept an argument, an external observer may still accept it. A dialogue is a sequence of steps from the framework to the extensions, where at each step an agent can commit to accept or reject some arguments, or commit to hide or reveal one of his/her rejected arguments. The revealed arguments are then aggregated and an external observer can compute which arguments are finally acceptable at a global level.

Our use of input/output argumentation as a model for multi-agent argumentation contains some assumptions, which can be relaxed in further research. In particular, the conditional reasoning of the agents implies that they do not have a model of the arguments pertaining to the other agents, and therefore the only agent who considers the interaction between agents is the external observer. In more sophisticated models, agents can recursively model other agents, including the other agents' model of their own arguments.

This article also considers the application of our framework to legal reasoning,

	Name	Components
Sec. 2 (Def. 2)	Argumentation framework	$\langle \mathcal{A}, \mathcal{R} \rangle$
Sec. 2 (Def. 4)	Multi-agent argumentation framework	$\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$
Sec. 3 (Def. 10)	Trust argumentation framework	$\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$
Sec. 3 (Def. 10)	Trust argumentation framework of an agent A	$\langle \mathcal{A}_A, \mathcal{R}_A, I_A, R_{I_A} \rangle$

Table 1: The different frameworks defined throughout the article.

using an example from a court case. Here, the agents are the accused, the lawyers, the witnesses, the prosecutors and so on. The external observer is the jury or the judge, who has to take into account all the arguments put forward during the deliberation and decide which arguments to accept. For example, we can distinguish the collective argumentation of judges from the individual argumentation of the accused, prosecutors, witnesses, lawyers, and experts. In multi-agent argumentation, agents have partial knowledge of the arguments and attacks of other agents, and they decide autonomously whether to accept or reject their own arguments as well as whether to bring their arguments forward in court. The arguments accepted by the judge are based on a game-theoretic equilibrium of the argumentation of the other agents. The multi-agent argumentation can be used to distinguish various direct and indirect ways in which an agent’s arguments can be used against his/her other arguments. The novelty of the framework we introduce in this article is that not only do we have agents with local argumentation frameworks and individual acceptance functions, but we also consider a social network for the agents, and the semantics also specify whether the agents decide to hide or reveal the arguments they do not accept.

Tables 1 and 2 provide a list of the concepts introduced in Sections 2, 3 and 4 of this article. Table 1 provides a list of the different structures introduced throughout the article, specifying where they are introduced and which components they are made of. Table 2 provides a list of the different acceptance functions defined in this article, together with a note of where they first appear. The second column is comprised of general acceptance functions, while the third column focuses on the local acceptance functions of each agent. Finally, the last column recounts the global collective acceptance functions that aggregate the individual acceptance functions.

The layout of this article is as follows. In Section 2, we repeat the definitions and concepts of the input/output argumentation of Baroni *et al.* [9], and we consider the limitations that arise when considering it as a kind of multi-agent argumentation. In

	Acceptance	Individual acceptance	Collective acceptance
Sec. 2	ac-arg (Def. 2)	iac-arg (Def. 6)	cac-arg (Def. 8)
Sec. 3	ac-trust (Def. 10)	iac-trust (Def. 10)	cac-trust (Def. 10)
Sec. 4	ac-sub (Def. 16)	iac-sub_A (Def. 17)	cac-sub (Def.18)

Table 2: The different acceptance functions defined throughout the article.

Section 3 we extend input/output argumentation with a social trust network, and in Section 4 we introduce the possibility of communicating arguments to other agents via sub-frameworks. In Section 5 we introduce dialogue semantics by refining the sub-framework approach with dialogue steps. In Section 6 we discuss related work.

2 Conditional and multi-agent argumentation

In this section, we introduce the basic definitions that pertain to multi-agent argumentation, including individual agents' conditional acceptance of arguments based on other agents' acceptance of the arguments. The formal theory of multi-agent argumentation in this section is an interpretation of so-called input/output argumentation as described by Baroni *et al.* [9]. Whereas input/output argumentation is a powerful theory for distinguishing the acceptance of arguments by individual agents from the acceptance of arguments by an external observer, we also explain why we need to extend input/output argumentation to make it applicable to multi-agent argumentation.

We first recall Dung's abstract argumentation semantics [26] which can be represented by a function associating sets of jointly acceptable arguments with argumentation frameworks. Though Dung introduced various ways of defining the semantics of argumentation frameworks, in this article we consider only so-called stable semantics in order to keep our formal exposition to a minimum.

Definition 1 (Stable semantics for argument acceptance). *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be a directed graph called an argumentation framework, where the elements of \mathcal{A} are called arguments and the binary relation \mathcal{R} over \mathcal{A} is called an attack relation. A subset of the arguments is a stable extension of the argumentation framework if and only if it does not contain an argument attacking another argument in the extension, and for each argument that is not in the extension, there is an argument in the extension attacking it. We write $\text{ac-arg}(\langle \mathcal{A}, \mathcal{R} \rangle)$ for the set of all stable extensions of the argumentation framework $\langle \mathcal{A}, \mathcal{R} \rangle$.*

The following example illustrates how Dung uses stable extensions of argumentation frameworks to define the concept of arguments that can be accepted together.

It also explains how there can be multiple extensions of an argumentation framework.¹ To find all stable extensions, one can guess an extension and check whether it is stable.

Example 2 (Four arguments). *Consider the arguments $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$ with attack relation $\mathcal{R} = \{(a_1, a_2), (a_2, a_1), (a_3, a_4), (a_4, a_3), (a_3, a_1), (a_2, a_4)\}$ visualised on the left-hand side of Figure 2. We have $\mathbf{ac}\text{-arg}(\langle \mathcal{A}, \mathcal{R} \rangle) = \{\{a_1, a_4\}, \{a_2, a_3\}\}$ since for $\{a_1, a_4\}$, neither a_1 nor a_4 is attacked by a_1 or a_4 , and the arguments attacking either of them, namely a_2 and a_3 , are attacked by them; analogous reasoning goes for $\{a_2, a_3\}$.*

In Dung’s semantics, acceptance firstly means no conflict. If we accept any argument, every argument attacked by it is thus certainly rejected. In a way, acceptance of an argument becomes the cause for rejecting other arguments. Each extension (member) of the stable semantics is stable in the sense that (1) no arguments in the extension are in conflict and (2) rejection of every other argument outside it is guaranteed.

Each extension in $\{\{a_1, a_4\}, \{a_2, a_3\}\}$ satisfies the two criteria. Moreover, no other sets of arguments are stable in this example. Any strict superset would involve a conflict. For instance, $\{a_1, a_3, a_4\}$ is in conflict since a_3 attacks a_1 . Any strict subset does not ensure rejection of all the other arguments. Consider $\{a_2\}$, while it rejects a_1 and a_4 , a_3 is not rejected by it.



Figure 2: An argumentation framework and a multi-agent argumentation framework.

A multi-agent argumentation framework is an argumentation framework with a set of agents and an assignment of the arguments to the agents. We also call the agents the *sources* of the arguments. Rienstra *et al.* [47] call it a multi-sorted argumentation framework, and Baroni *et al.* [9] call it an input/output argumentation framework.

Definition 2 (Multi-agent argumentation). *A multi-agent argumentation framework is a tuple $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ where $\langle \mathcal{A}, \mathcal{R} \rangle$ is an argumentation framework, Ag is*

¹In fact, there can even be argumentation frameworks that do not have any stable extensions, for example argumentation frameworks consisting of a single argument attacking itself. This reflects the notion of incoherence in argumentation.

a set called agents and $Src : \mathcal{A} \rightarrow Ag$ is a function mapping each argument to the agent that put it forward (also known as its source).

Example 3 (Two agents, continued from Example 2). Consider the multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ visualised on the right-hand side of Figure 2. We have $Ag = \{W_1, W_2\}$, $Src(a_1) = Src(a_2) = W_1$, and $Src(a_3) = Src(a_4) = W_2$.

For the semantics, we first define individual acceptance by an agent. We consider the part of the multi-agent framework that is relevant to the agent, which we call the *agent argumentation framework*. It contains its own arguments together with the attacks from and against those arguments, the relevant arguments of other agents, an extension of these other arguments, and an attack relation from these other arguments to its own arguments. The agent semantics considers the agent argumentation framework as well as the arguments accepted by other agents. This conditional acceptance is called a local function by Baroni *et al.* [9].

We slightly rewrite the definition of local function to make it explicit that agents' acceptance of arguments is conditional on the other agents' accepted arguments. Moreover, in contrast to Baroni *et al.* [9], we do not consider attacks against input arguments. Since Baroni *et al.* [9] define their local acceptance functions for all Dung semantics and not only for stable semantics, their definitions are more general than ours. Similar notions are defined also by Liao [38]. We refer to these papers for further explanations and examples of local functions.

Definition 3 (Individual conditional acceptance). For multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$, the argumentation framework of agent A is a tuple $\langle \mathcal{A}_A, \mathcal{R}_A, I_A, R_{I_A} \rangle$, where $\mathcal{A}_A = \{a \in \mathcal{A} \mid Src(a) = A\}$ are the arguments of agent A , $\mathcal{R}_A = \mathcal{R} \cap (\mathcal{A}_A \times \mathcal{A}_A)$ are its attacks, $I_A = \{a \in \mathcal{A} \mid a \notin \mathcal{A}_A, (a, b) \in \mathcal{R}, b \in \mathcal{A}_A\}$ are the relevant arguments from other agents, and $R_{I_A} = \mathcal{R} \cap (I_A \times \mathcal{A}_A)$ is the corresponding attack relation. The stable semantics of agent A and context $E_{I_A} \subseteq I_A$, a set of arguments called the input extension, is defined by

$$iac\text{-}arg(\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle, A, E_{I_A}) = \\ ac\text{-}arg(\langle \mathcal{A}_A \cup E_{I_A}, \mathcal{R}_A \cup (\mathcal{R} \cap (E_{I_A} \times \mathcal{A}_A)) \rangle)_{\mathcal{A}_A}$$

where for a set of extensions S , $S_{\mathcal{A}_A} = \{s \cap \mathcal{A}_A \mid s \in S\}$. We may denote a member of $iac\text{-}arg(\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle, A, E_{I_A})$ by E_A .

Conditional acceptance is illustrated in the example below.

Example 4 (Two agents, continued from Example 3). Figure 3 visualises the agent argumentation frameworks from the running example. The left-hand side visualises

the agent argumentation framework of agent W_1 , and the right-hand side visualises the framework of agent W_2 . The left framework represents a tuple $\langle \mathcal{A}_{W_1}, \mathcal{R}_{W_1}, I_{W_1}, R_{I_{W_1}} \rangle$ where $\mathcal{A}_{W_1} = \{a_1, a_2\}$ are the arguments of agent W_1 , $\mathcal{R}_{W_1} = \{(a_1, a_2), (a_2, a_1)\}$ are its attacks, $I_A = \{a_3\}$ are the relevant arguments from other agents, and $R_{I_A} = \{(a_3, a_1)\}$ is the corresponding attack relation.



Figure 3: The two agent frameworks from the multi-agent framework in Figure 2.

The stable semantics of agent W_1 depends not only on its agent framework, but also on its input extension. In other words, agent W_1 's acceptance of arguments is conditional on its input extension, which is why we refer to the semantics of individual agents as conditional semantics. This input extension of agent W_1 is a subset of its input arguments $\{a_3\}$, so the input extension is either the empty set or $\{a_3\}$. The stable semantics for agent W_1 is now calculated to either reject or accept a_3 .

In the former case, the agent accepts either a_1 or a_2 , and in the latter case, the agent accepts a_2 . We have $\text{iac-arg}(\langle \mathcal{A}_{W_1}, \mathcal{R}_{W_1}, I_{W_1}, R_{I_{W_1}} \rangle, W_1, \emptyset) = \{\{a_1\}, \{a_2\}\}$ and $\text{iac-arg}(\langle \mathcal{A}_{W_1}, \mathcal{R}_{W_1}, I_{W_1}, R_{I_{W_1}} \rangle, W_1, \{a_3\}) = \{\{a_2\}\}$.

Likewise, for agent W_2 , the input contains only argument a_2 from agent W_1 , and the input extension is either the empty set or $\{a_2\}$. In the former case, agent W_2 accepts either argument a_3 or a_4 , and in the latter case, the agent accepts a_3 . We have $\text{iac-arg}(\langle \mathcal{A}_{W_2}, \mathcal{R}_{W_2}, I_{W_2}, R_{I_{W_2}} \rangle, W_2, \emptyset) = \{\{a_3\}, \{a_4\}\}$ and also $\text{iac-arg}(\langle \mathcal{A}_{W_2}, \mathcal{R}_{W_2}, I_{W_2}, R_{I_{W_2}} \rangle, W_2, \{a_2\}) = \{\{a_3\}\}$.

We finally provide a definition for collective acceptance, which may be seen as the arguments accepted by an external observer.

Definition 4 (Collective acceptance). *The collective stable semantics of multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$, which we write as $\text{cac-arg}(\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle)$, is the set of extensions $S \subseteq \mathcal{A}$ such that for all agents $A \in Ag$ we have*

$$S \cap \mathcal{A}_A \in \text{iac-arg}(\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle, A, S \cap I_A).$$

The following example illustrates how to check whether an extension is collectively accepted.

Example 5 (Two agents, continued from Example 4). *An external observer now has to combine the two semantic functions of agents W_1 and W_2 to find the collectively accepted arguments. Since each agent has only three possibilities, there are only nine cases to check. As it turns out, only two of them are compatible and thus lead to a collective extension. The stable extensions are $\text{cac-arg}(\langle \mathcal{A}, \mathcal{R}, \text{Ag}, \text{Src} \rangle) = \{\{a_1, a_4\}, \{a_2, a_3\}\}$.*

	$E_{I_{W_1}} = \emptyset,$ $E_{W_1} = \{a_1\}$	$E_{I_{W_1}} = \emptyset,$ $E_{W_1} = \{a_2\}$	$E_{I_{W_1}} = \{a_3\},$ $E_{W_1} = \{a_2\}$
$E_{I_{W_2}} = \emptyset, E_{W_2} = \{a_3\}$	x	x	x
$E_{I_{W_2}} = \emptyset, E_{W_2} = \{a_4\}$	$\{a_1, a_4\}$	x	x
$E_{I_{W_2}} = \{a_2\}, E_{W_2} = \{a_3\}$	x	x	$\{a_2, a_3\}$

To compute the extensions, one can guess an extension and then check that it is in equilibrium in the sense that for every agent, if the input is part of the extension, then the agent accepts the arguments provided by the individual acceptance function.

The reader may observe that the two extensions of the collective semantics coincide with the two stable extensions of the argumentation framework in Example 2, and wonder whether this holds more generally. Perhaps surprisingly, Baroni *et al.* [9] prove that this is no coincidence. For many semantics σ , including stable semantics, collective acceptance coincides with the σ of the underlying argumentation framework:

$$\text{cac-arg}(\langle \mathcal{A}, \mathcal{R}, \text{Ag}, \text{Src} \rangle) = \text{ac-arg}(\langle \mathcal{A}, \mathcal{R} \rangle)$$

In their approach, this represents a useful principle, because it allows for compositional computation of the semantics, and various applications such as summarisation. As for using input/output argumentation as a theory of multi-agent argumentation, it shows that the theory must be extended. For instance, Rienstra *et al.* [47] study the case where the individual acceptance of the agents (sorts in their terminology) may use different semantics. As an example, one agent may use stable semantics as described in this article, while another agent may use grounded or preferred semantics. In this article, we extend the theory in other ways by introducing the notion of trust in Section 3, allowing agents to hide information in Section 4, and then considering communication between agents in Section 5.

3 Multi-agent argumentation with a social trust network

In this section, we extend multi-agent argumentation with a social network for agents reflecting epistemic trust. An agent trusts another agent if the first agent accepts the arguments the second agent accepts. If the social network is reflexive, symmetric and transitive, then the network consists of equivalence classes of agents, which may be called coalitions.

Individual and collective acceptance in trust argumentation frameworks are defined the same as before, using trust argumentation frameworks for individual agents.

Definition 5 (Trust argumentation framework). *A trust argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$ extends a multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ with a binary relation of trust $T \subseteq Ag \times Ag$ such that each agent A trusts itself, i.e. $(A, A) \in T$, which we can write alternatively as $T(A, A)$. We write $T(A)$ for $\{B \mid T(A, B)\}$.*

For trust argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$, we say that $\langle \mathcal{A}, \mathcal{R}', Ag, Src \rangle$ with $\mathcal{R}' = \{(a, b) \in \mathcal{R} \mid T(Src(b), Src(a))\}$ is its Multi-agent Argumentation Framework (MAF) representation.

$$\begin{aligned} iac\text{-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle, A, E_{I_A}) &= \\ iac\text{-arg}(\langle \mathcal{A}, \mathcal{R}', Ag, Src \rangle, A, E_{I_A}) & \\ cac\text{-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle) &= cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}', Ag, Src \rangle) \end{aligned}$$

Example 6 (Two agents, continued from Example 5). *Let us consider the multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ on the right-hand side of Figure 2, and let us extend it with trust relation T to derive $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$. Figure 4 visualises its MAF representations for each choice of T .*

- $T(W_1) = T(W_2) = \{W_1, W_2\}$ means that the agents trust each other. The top-left corner of Figure 4 shows the corresponding MAF representation.
 $cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_1, \{a_3\}) = \{\{a_2\}\}$.
 $cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_1, \emptyset) = \{\{a_1\}, \{a_2\}\}$.
 $cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_2, \{a_2\}) = \{\{a_3\}\}$.
 $cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_2, \emptyset) = \{\{a_3\}, \{a_4\}\}$.
 $cac\text{-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle) = \{\{a_1, a_4\}, \{a_2, a_3\}\}$.
- $T(W_1) = \{W_1, W_2\}$, $T(W_2) = \{W_2\}$ means that only agent W_1 trusts agent W_2 . The top-right corner of Figure 4 shows the corresponding MAF representation.

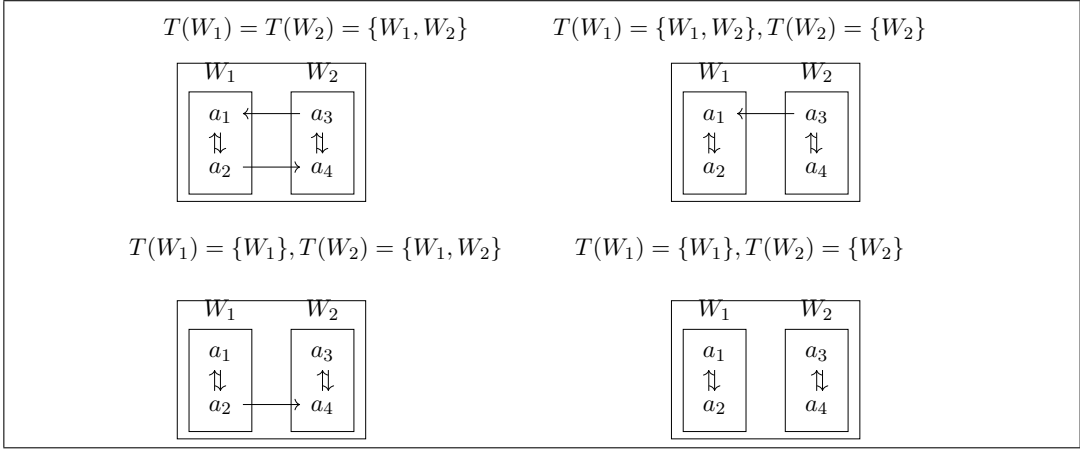


Figure 4: MAF representations for a chosen trust relation.

$$\begin{aligned}
cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_1, \{a_3\}) &= \{\{a_2\}\}. \\
cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_1, \emptyset) &= \{\{a_1\}, \{a_2\}\}. \\
cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_2, \emptyset) &= \{\{a_3\}, \{a_4\}\}. \\
cac\text{-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle) &= \{\{a_1, a_4\}, \{a_2, a_3\}, \{a_2, a_4\}\}.
\end{aligned}$$

- $T(W_1) = \{W_1\}$, $T(W_2) = \{W_1, W_2\}$ means that only agent W_2 trusts agent W_1 . The bottom-left corner of Figure 4 shows the corresponding MAF representation.

$$\begin{aligned}
cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_1, \emptyset) &= \{\{a_1\}, \{a_2\}\}. \\
cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_2, \{a_2\}) &= \{\{a_3\}\}. \\
cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_2, \emptyset) &= \{\{a_3\}, \{a_4\}\}. \\
cac\text{-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle) &= \{\{a_1, a_3\}, \{a_1, a_4\}, \{a_2, a_3\}\}.
\end{aligned}$$

- $T(W_1) = \{W_1\}$, $T(W_2) = \{W_2\}$ means the agents do not trust one other. The bottom-right corner of Figure 4 shows the corresponding MAF representation.

$$\begin{aligned}
cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_1, \emptyset) &= \{\{a_1\}, \{a_2\}\}. \\
cac\text{-arg}(\langle \mathcal{A}, \mathcal{R}, Ar, Src, T \rangle, W_2, \emptyset) &= \{\{a_3\}, \{a_4\}\}. \\
cac\text{-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle) &= \{\{a_1, a_3\}, \{a_1, a_4\}, \{a_2, a_3\}, \{a_2, a_4\}\}.
\end{aligned}$$

In the above example, the lack of trust leads to an increase in the number of extensions in $cac\text{-trust}$. The following example illustrates that this is not always the case.

Example 7 (Two agents, another example). Consider the trust argumentation framework in Figure 5. Let us extend it with a trust relation. Depending on which

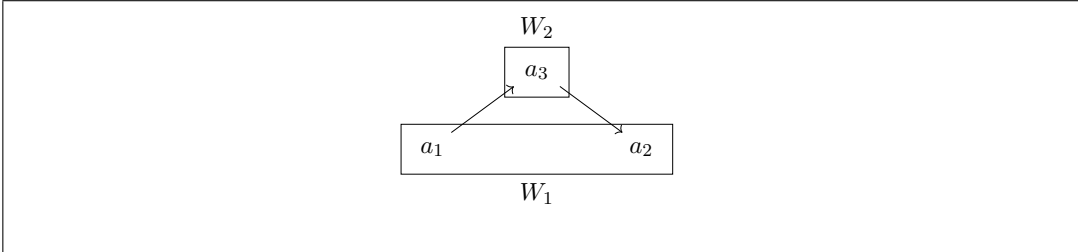


Figure 5: Multi-agent argumentation with 2 agents, W_1 and W_2 .

agents trust whom, we obtain four different MAF representations as shown in Figure 6.

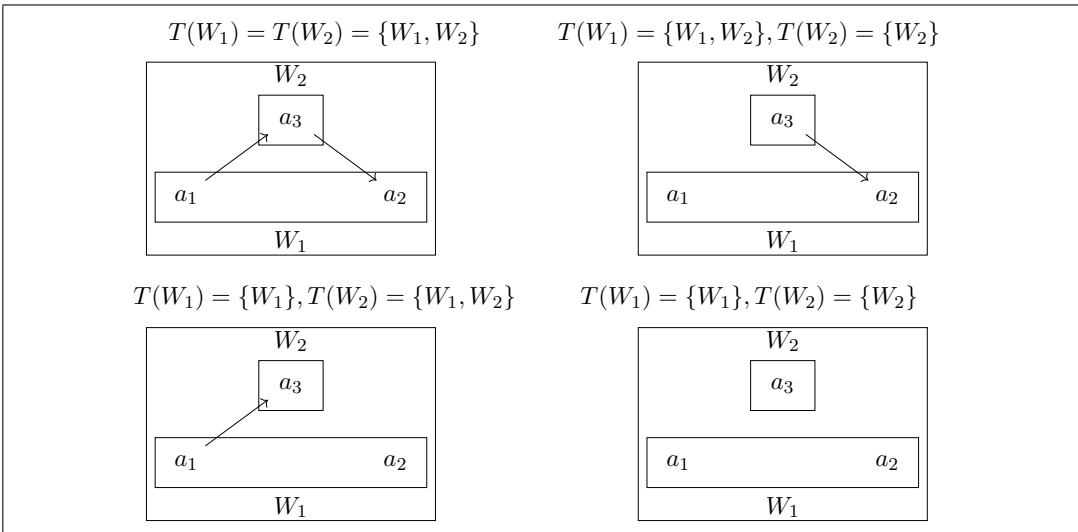


Figure 6: MAF representations for a chosen trust relation.

- $T(W_1) = T(W_2) = \{W_1, W_2\}$ means that the agents trust each other. The top-left corner of Figure 6 shows the corresponding MAF representation. $\text{cac-trust}(\langle \mathcal{A}, \mathcal{R}, \text{Ag}, \text{Src}, T \rangle) = \{\{a_1, a_2\}\}$.
- $T(W_1) = \{W_1, W_2\}, T(W_2) = \{W_2\}$ means that only W_1 trusts the other. The top-right corner of Figure 6 shows the corresponding MAF representation. $\text{cac-trust}(\langle \mathcal{A}, \mathcal{R}, \text{Ag}, \text{Src}, T \rangle) = \{\{a_1, a_3\}\}$.
- $T(W_2) = \{W_1, W_2\}, T(W_1) = \{W_1\}$ means that only W_2 trusts the other. The bottom-left corner of Figure 6 shows the corresponding MAF representation.

$$\text{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle) = \{\{a_1, a_2\}\}.$$

- $T(W_1) = \{W_1\}, T(W_2) = \{W_2\}$ means that each agent trusts only himself/herself. The bottom-right corner of Figure 6 shows the corresponding MAF representation.

$$\text{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle) = \{\{a_1, a_2, a_3\}\}.$$

As is clear from these examples, it is not necessary that collective acceptance of a trust argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$ should consist solely of conflict-free extensions (*i.e.* those extensions in which no attacks occur), unlike in the multi-agent argumentation frameworks in the previous section. When we have an agent W in a trust argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$, and when, to ensure that W 's arguments do not cause a conflict, collective acceptance of any of W 's arguments necessarily implies non-acceptance of arguments put forward by other agents attacking it, it is in our interest to know what must be minimally added to this trust argumentation framework in order to break the no-conflict property of collective acceptance. By establishing the knowledge, we can consecutively learn what additions to the trust argumentation framework are safe to make without violating the no-conflict property.

It turns out that if an extension S'' in $\text{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle)$ includes some agent W 's argument a_1 and another agent W' 's argument a_2 , then $\langle \mathcal{A}, \mathcal{R} \cup \{(a_2, a_1)\}, Ag, Src, T \rangle$ already breaks the no-conflict property provided that (1) W does not trust W' , and provided also that (2) there is some extension S in $\text{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle)$ and some extension S' in $\text{cac-trust}(\langle \mathcal{A}, \mathcal{R} \cup \{(a_2, a_1)\}, Ag, Src, T \rangle)$ such that S includes a_1 and a_2 , and that S' contains a_2 and any $a_x \in (S \setminus \mathcal{A}_W)$ attacking an argument in \mathcal{A}_W but no other $a_y \in (\mathcal{A} \setminus \mathcal{A}_W)$ attacking an argument in \mathcal{A}_W . Furthermore, the opposite is also true. If either of the conditions (1) or (2) above is not satisfied, then no extension in $\text{cac-trust}(\langle \mathcal{A}, \mathcal{R} \cup \{(a_2, a_1)\}, Ag, Src, T \rangle)$ contains both a_1 and a_2 .

We make use of the following two lemmas to prove the result.

Lemma 1. *For any $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$, any $W_1, W_2 \in Ag$, any $a_1 \in \mathcal{A}_{W_1}$, and any $a_2 \in \mathcal{A}_{W_2}$, if $(a_2, a_1) \in \mathcal{R}$ and if there is some $S \in \text{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle)$ such that $a_1, a_2 \in S$, then $W_2 \notin T(W_1)$.*

Proof. We show the contrapositive. Let $\langle \mathcal{A}, \mathcal{R}', Ag, Src \rangle$ be its MAF representation. If $W_2 \in T(W_1)$, then for any $a \in \mathcal{A}_{W_1}$, $(a_2, a_1) \in \mathcal{R}'$ holds iff $(a_2, a_1) \in \mathcal{R}$ holds. By the definition of iac-trust , for all $S_{W_1} \in \text{iac-trust}(\langle \mathcal{A}_{W_1}, \mathcal{R}_{W_1}, I_{W_1}, R_{I_{W_1}} \rangle, W_1, \{a_2\})$, it holds that $a_1 \notin S_{W_1}$. Then, for any $S \in \text{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle)$, if

$a_2 \in S$, it must be the case that $a_1 \notin S$; and by the contrapositive, if $a_1 \in S$, it must also be the case that $a_2 \notin S$, as required. \square

Lemma 2. *For any $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$, any $W \in Ag$, and any $a_1 \in \mathcal{A}_W$, if there is some $S \in \mathbf{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle)$ such that $a_1 \in S$, then for any $\langle \mathcal{A}, \mathcal{R}^*, Ag, Src, T \rangle$ with $\mathcal{R} \setminus (\bigcup_{W_x \in (Ag \setminus \{W\})} \mathcal{A}_{W_x} \times \{a_1\}) \subseteq \mathcal{R}^* \subseteq \mathcal{R}$, there is some $S' \in \mathbf{cac-trust}(\langle \mathcal{A}, \mathcal{R}^*, Ag, Src, T \rangle)$ such that $S_W = S'_W$.*

Proof. Let $\langle \mathcal{A}, \mathcal{R}', Ag, Src \rangle$ be the MAF representation of $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$, let $\langle \mathcal{A}, \mathcal{R}'', Ag, Src \rangle$ be the MAF representation of $\langle \mathcal{A}, \mathcal{R}^*, Ag, Src, T \rangle$, let I'_W denote $\{a \in \mathcal{A} \setminus \mathcal{A}_W \mid \exists a_x \in \mathcal{A}_W. (a, a_x) \in \mathcal{R}'\}$, and let I''_W denote $\{a \in \mathcal{A} \setminus \mathcal{A}_W \mid \exists a_x \in \mathcal{A}_W. (a, a_x) \in \mathcal{R}''\}$. For any $E_{I'_W} \subseteq I'_W$, if $S_W \in \mathbf{iac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle, W, E_{I'_W})$, then for any $a_x \in (\mathcal{A} \setminus \mathcal{A}_W)$, if $(a_x, a_1) \in \mathcal{R}$, then $a_x \notin E_{I'_W}$, because no stable extension of $(\mathcal{A}_W \cup E_{I'_W}, \mathcal{R}_W \cup R_{E_{I'_W}})$ with $R_{E_{I'_W}} \equiv \{(a_2, a_1) \in \mathcal{R}' \mid a_2 \in E_{I'_W} \text{ and } a_1 \in \mathcal{A}_W\}$ contains two arguments with one (or both) attacking another. But this then implies that $S \cap I'_W = S \cap I''_W$. Meanwhile, for any $W_x \in (Ag \setminus \{W\})$, it trivially holds that $I'_{W_x} = I''_{W_x}$. Hence, for any $W_y \in Ag$, we have $S \cap I'_{W_y} = S \cap I''_{W_y}$, so $\mathbf{iac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle, W_y, S \cap I'_{W_y}) = \mathbf{iac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle, W_y, S \cap I''_{W_y})$. It then holds that $S \in \mathbf{cac-trust}(\langle \mathcal{A}, \mathcal{R}^*, Ag, Src, T \rangle)$. Let S' thus denote S , to conclude. \square

Theorem 1 (Collective acceptance with trust). *For any $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$ and any $W \in Ag$, let I_a for $a \in \mathcal{A}$ denote $\{a_3 \in \mathcal{A} \mid Src(a_3) \neq Src(a) \text{ and } (a_3, a) \in \mathcal{R}\}$. Given a $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$ and a $W \in Ag$, if $(\bigcup_{a \in S_W} I_a) \cap S = \emptyset$ for any $S \in \mathbf{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle)$, then for any $a_1 \in \mathcal{A}_W$, any $a_2 \notin \mathcal{A}_W$ and any $\langle \mathcal{A}, \mathcal{R} \cup \{(a_2, a_1)\}, Ag, Src, T \rangle$, the first condition below holds just when the second condition holds.*

1. (1) *There is some $S \in \mathbf{cac-trust}(\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle)$ and some $S' \in \mathbf{cac-trust}(\langle \mathcal{A}, \mathcal{R} \cup \{(a_2, a_1)\}, Ag, Src, T \rangle)$ such that $a_1, a_2 \in S$, and that $S \cap (\{a_2\} \cup \bigcup_{a \in \mathcal{A}_W} I_a) = S' \cap \bigcup_{a \in \mathcal{A}_W} I_a$. (2) $Src(a_2) \notin T(W)$.*
2. *There is some $S'' \in \mathbf{cac-trust}(\langle \mathcal{A}, \mathcal{R} \cup \{(a_2, a_1)\}, Ag, Src, T \rangle)$ such that $a_1, a_2 \in S''$.*

Proof. Let $\langle \mathcal{A}, \mathcal{R}', Ag, Src \rangle$ be the MAF representation of $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$, let $\langle \mathcal{A}, \mathcal{R}'', Ag, Src \rangle$ be the MAF representation of $\langle \mathcal{A}, \mathcal{R} \cup \{(a_2, a_1)\}, Ag, Src, T \rangle$, let I'_W denote $\{a \in \mathcal{A} \setminus \mathcal{A}_W \mid \exists a_x \in \mathcal{A}_W. (a, a_x) \in \mathcal{R}'\}$, and let I''_W denote $\{a \in \mathcal{A} \setminus \mathcal{A}_W \mid \exists a_x \in \mathcal{A}_W. (a, a_x) \in \mathcal{R}''\}$.

1 to 2: Due to (2), it holds that $\langle \mathcal{A}, \mathcal{R}' \rangle = \langle \mathcal{A}, \mathcal{R}'' \rangle$. Thus, for any $W_x \in Ag$, we

have $I'_{W_x} = I''_{W_x}$. The reasoning for the rest is similar to that provided in the proof for Lemma 2.

2 to 1: (2) follows from Lemma 1. The first part of (1), i.e. that some such S includes a_1 and a_2 , follows from the assumption on $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$. The second part of (1) is via Lemma 2. \square

We can also refine the trust relation in terms of topics of expertise, such that an agent only accepts an argument put forward by another agent if it concerns a topic on which the first agent trusts the second agent to be informed.

4 Communication of arguments

In this section, we further extend multi-agent argumentation by allowing agents to communicate not only the arguments they accept, but also some of the arguments they reject, and we describe how they all interact.

The following definition generalises Dung's stable extensions as sub-frameworks. A stable sub-framework is a sub-framework having exactly one stable extension that also is a stable extension of the whole framework. A sub-framework semantics called AFRA (argumentation framework with recursive attacks) semantics was introduced by Baroni *et al.* [10] and a sub-framework semantics called attack semantics was introduced by Villata *et al.* [58].

Definition 6 (Stable sub-frameworks). *The framework $\langle \mathcal{A}', \mathcal{R}' \rangle$ is a stable sub-framework of $\langle \mathcal{A}, \mathcal{R} \rangle$ if and only if $\mathcal{A}' \subseteq \mathcal{A}$, $\mathcal{R}' \subseteq \mathcal{R} \cap (\mathcal{A}' \times \mathcal{A}')$ and $\langle \mathcal{A}', \mathcal{R}' \rangle$ has exactly one stable extension which is also a stable extension of $\langle \mathcal{A}, \mathcal{R} \rangle$. We write $\mathbf{ac-sub}(\langle \mathcal{A}, \mathcal{R} \rangle)$ for the set of all stable sub-frameworks of the argumentation framework $\langle \mathcal{A}, \mathcal{R} \rangle$.*

Every Dung extension can be interpreted as a sub-framework containing only the accepted arguments and an empty attack relation. However, in general, there are several sub-frameworks corresponding to each Dung extension. We use this fact to define an agent's individual acceptance function $\mathbf{ac-sub}_A$. For example, $\mathbf{ac-sub}_A$ can be the minimal sub-framework corresponding to Dung's extension, but it can also be a maximal sub-framework communicating as much information as possible.

Given a $\mathbf{ac-sub}_A$, the definition of an agent's individual acceptance function stays exactly the same, except it replaces $\mathbf{ac-arg}$ with $\mathbf{ac-sub}_A$.

Definition 7 (Individual acceptance). *For a multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ with agent $A \in Ag$, an individual acceptance function $\mathbf{ac-sub}_A$*

returns a set of stable sub-frameworks $\mathbf{ac-sub}_A(\langle \mathcal{A}, \mathcal{R} \rangle) \subseteq \mathbf{ac-sub}(\langle \mathcal{A}, \mathcal{R} \rangle)$ containing at least one sub-framework for each stable extension, so $\bigcup_{F \in \mathbf{ac-sub}_A(\langle \mathcal{A}, \mathcal{R} \rangle)} \mathbf{ac-arg}(F) \supseteq \mathbf{ac-arg}(\langle \mathcal{A}, \mathcal{R} \rangle)$.

For a multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ and an individual acceptance function $\mathbf{ac-sub}_A$, the stable semantics of agent A and context $E_{I_A} \subseteq I_A$ is defined by

$$\begin{aligned} \mathbf{iac-sub}_A(\langle \mathcal{A}_A, \mathcal{R}_A, I_A, R_{I_A} \rangle, E_{I_A}, \mathbf{ac-sub}_A) = \\ \mathbf{ac-sub}_A(\langle \mathcal{A}_A \cup E_{I_A}, \mathcal{R}_A \cup (R_{I_A} \cap (E_{I_A} \times \mathcal{A}_A)) \rangle)_{\mathcal{A}_A}. \end{aligned}$$

Moreover, the definition of collective acceptance is adapted as follows.

Definition 8 (Collective acceptance). *A stable sub-framework of multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ with the set $\{\mathbf{ac-sub}_A \mid A \in Ag\}$ of individual acceptance functions is a triple $\langle \mathcal{A}', \mathcal{R}', S \rangle$ such that S is a stable extension of $\langle \mathcal{A}', \mathcal{R}' \rangle$ and for all agents $A \in Ag$, we have*

1. $\langle \mathcal{A}', \mathcal{R}' \rangle_A \in \mathbf{ac-sub}_A(\langle \mathcal{A}_A, \mathcal{R}_A, I_A, R_{I_A} \rangle, S \cap E_{I_A})_A$, and
2. for all (a_1, a_2) such that $Src(a_1) \neq Src(a_2)$, we have $(a_1, a_2) \in \mathcal{R}'$ iff $(a_1, a_2) \in \mathcal{R} \cap (\mathcal{A}' \times \mathcal{A}')$.

We write $\mathbf{cac-sub}(\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle, \{\mathbf{ac-sub}_A \mid A \in Ag\})$ for the set of all such triples from multi-agent argumentation framework $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ with individual acceptance functions $\{\mathbf{ac-sub}_A \mid A \in Ag\}$.

The first item ensures that the collectively accepted framework locally corresponds to each agent's accepted sub-framework, while the second item states that attacks from one agent's individually accepted framework on another are as they are in the original framework. Trust networks and sub-framework semantics can be combined in the obvious way.

Example 8 below illustrates a multi-agent debate involving an accused, a witness, a prosecutor and finally a judge who is evaluating collective acceptance. We show how variations in individual conditional acceptance, i.e. variations in what to reveal for the judgment of collective acceptance, can lead to different outcomes, some good and some bad for the accused.

Example 8. *There was a murder at Laboratory C which Acc is accused of having committed. There is a witness Wit and a prosecutor Prc. Acc has two arguments:*

a_1 : *that he was at Laboratory A on the day of the murder (this is a fact known to Acc).*

a_2 : that he is innocent (this is *Acc*'s claim).

Prc entertains:

a_6 : that only *Acc* could have killed the victim (this is *Prc*'s claim).

Meanwhile, *Wit* believes certain things. He has three arguments:

a_3 : *Acc* stayed at home on the day of the murder, having previously lost his ID card
(*Wit* originally believes this to be a fact).

a_4 : *Acc* could enter any laboratory provided that he does so with his own ID card
(this is a fact known to *Wit*).

a_5 : *Acc* could not have been at Laboratory C at the time of the murder (this is
Wit's claim).

The multi-agent debate in Figure 7 (A) represents this example, showing the conflicts between the arguments. We denote this multi-agent argumentation framework by $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$.

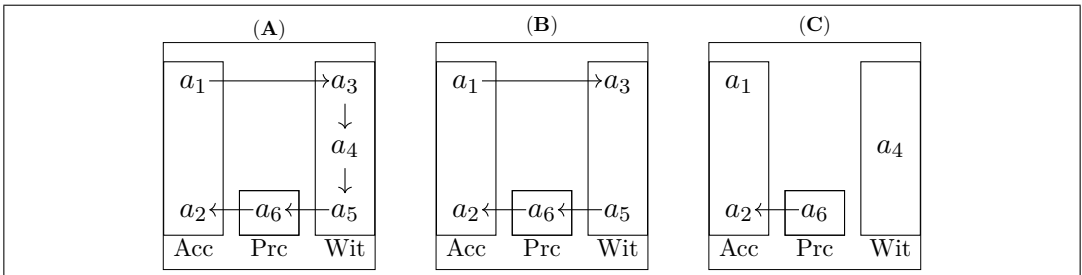


Figure 7: Accused (*Acc*), witness (*Wit*), and prosecutor (*Prc*).

In this example, *Prc* has no reason to drop her argument a_6 . Neither does *Acc*, seeing no benefit in conceding to a_6 , have any reason to drop a_2 . Hence, we only consider contexts where $E_{I_{Acc}} = E_{I_{Prc}} = \emptyset$. How *Wit* responds to the fact known to *Acc* (a_1), however, can prove crucial for *Acc* to be judged innocent or guilty by the judge who computes collective acceptance.

Case A. Suppose that $E_{I_{Wit}} = \emptyset$, which signifies that *Wit* is either not aware of a_1 or just ignores it. Then any individual acceptance function of *Wit*'s will output a set of sub-frameworks of $\langle \mathcal{A}, \mathcal{R} \rangle_{Wit}$ such that it has $\{a_3, a_5\}$ as its one and only stable extension. Thus, suppose that the acceptance function outputs only $\langle \mathcal{A}, \mathcal{R} \rangle_{Wit}$,

as shown in Figure 7 (A). Then, the stable sub-framework of $\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle$ is $\langle \mathcal{A}, \mathcal{R}, \{a_1, a_4, a_6\} \rangle$. Hence *Acc* is not judged innocent by the judge.

On the other hand, if *Wit*'s acceptance function outputs only $\langle \{a_3, a_5\}, \emptyset \rangle$, as shown in Figure 7 (B), then the stable sub-framework is $\langle \mathcal{A} \setminus \{a_4\}, \mathcal{R} \setminus (\{a_4\} \times \mathcal{A} \cup \mathcal{A} \times \{a_4\}), \{a_1, a_2, a_5\} \rangle$. *Acc* is judged innocent by the judge.

Case B. Suppose that $E_{I_{Wit}} = \{a_1\}$, which signifies that *Wit* takes a_1 into account. Then any individual acceptance function of *Wit*'s will output a set of sub-frameworks of $\langle \mathcal{A}, \mathcal{R} \rangle_{Wit}$ such that it has $\{a_4\}$ as its one and only stable extension. Thus, suppose that the acceptance function outputs only $\langle \{a_4\}, \emptyset \rangle$, as shown in Figure 7 (C). Then the stable sub-framework is $\langle \mathcal{A} \setminus \{a_3, a_5\}, \mathcal{R} \setminus (\{a_3, a_5\} \times \mathcal{A} \cup \mathcal{A} \times \{a_3, a_5\}), \{a_1, a_4, a_6\} \rangle$. Again, *Acc* is not judged innocent by the judge.

5 Dialogue semantics

In this section, we consider dialogical and dynamic aspects of multi-agent argumentation. Based on the work of Dauphin *et al.* [23] that refines extensions into decision graphs, we observe the process of argument sharing between agents and how the individual attitude of each agent affects the global outcome of the argumentation process.

We apply the same commitment-graph structure to argumentation dialogue between agents. The agents first commit to a single extension to their internal framework when multiple frameworks exist, and then decide how to share it. They may opt to fully share their arguments, exposing counter-arguments that they know about but locally reject, or they may decide to share the arguments they accept without mentioning any of the counter-arguments they are aware of. We represent these strategic choices in a graph to have a clearer visualisation of the communication process between different agents. We later provide a way to reduce these graphs such that only the choices that impact the final outcome are displayed, thus providing a summary of the argumentation process.

In their original work, Dauphin *et al.* [23] provided a framework for analysing the decisions made in the process of selecting one extension from a set of extensions. The decisions are represented in directed graphs where the nodes represent commitments made by the agent towards a progressively smaller subset of the set of extensions until only one extension remains.

We slightly adapt this framework when applying this approach to our multi-agent dialogue setting. When it is their turn, agents either simply decide which arguments to accept, or, once that is done, they also choose which arguments to share. While they will want to share every argument that they accept, that may not

be the case for the arguments they reject. For each argument they are aware of but do not accept, agents have the opportunity to either leave them out or share them with their peers.

We first present a definition of commitments about arguments. We allow our agents to choose which arguments to accept or reject, and for the arguments they reject, to choose whether or not to communicate this. We therefore introduce pairs to represent these decisions.

Definition 9. *Given $c \in \{+, -, \text{say}, \text{hide}\}$ and an argument a , we say that a pair (c, a) is a commitment on a . Given a set C of commitments on arguments, we say that C is coherent if there is no argument a such that:*

- $(\text{say}, a) \in C$ and $(\text{hide}, a) \in C$, or
- $(+, a) \in C$ and $(-, a) \in C$, or
- $(\text{hide}, a) \in C$ and $(+, a) \in C$.

For the sake of simplicity, we write $s(a)$ instead of (say, a) , we write $h(a)$ instead of (hide, a) , we write $+(a)$ instead of $(+, a)$, we write $-(a)$ instead of $(-, a)$. And for a set of commitments C , we write C^s for $\{a \mid s(a) \in C\}$, we write C^h for $\{a \mid h(a) \in C\}$, we write C^+ for $\{a \mid +(a) \in C\}$ and we write C^- for $\{a \mid -(a) \in C\}$.

We then define a sub-framework acceptance semantics that takes these commitments into account.

Definition 10. *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation framework and C a coherent set of commitments on arguments in \mathcal{A} . We define the C -committed stable sub-framework semantics to be*

$$\text{ac-com}^C(\langle \mathcal{A}, \mathcal{R} \rangle) = \{ \langle \mathcal{A} \setminus C^h, (\mathcal{R} \cap (\mathcal{A} \setminus C^h)^2) \setminus \mathcal{A} \times \mathcal{E} \mid \mathcal{E} \in \text{ac-arg}(\langle \mathcal{A}, \mathcal{R}, \mathcal{E} \supseteq C^+ \rangle) \}$$

The C -committed stable semantics first removes all the arguments that an agent has committed to hide. It then looks at the remaining framework, and for each stable extension containing the arguments the agent has committed to accept, it proceeds to remove the attacks on any arguments in that stable extension. This then forces one stable extension to apply to each framework from the output.

We can now adapt the decision graph structure in [23] to the triple-A frameworks. The individual agents still have to commit to which arguments they accept when their internal argumentation allows for multiple extensions, but then they also have to commit to which ones they communicate. We have defined the notion of C -committed stable semantics in order to let agents choose, for each argument they

reject, whether or not to share it. We now examine the impact these commitments have on the final extensions determined by the overall observer, or in our running examples, the judge.

Definition 11. Let $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$ be a trust argumentation framework. We say that a labelled directed acyclic graph $(\mathcal{C}, \mathcal{V}, l)$, where \mathcal{C} is a set of sets of commitments about \mathcal{A} , \mathcal{V} is a relation between elements of \mathcal{C} and l is a function assigning labels to both \mathcal{C} and \mathcal{V} (as described below), is a multi-agent commitment graph for $\langle \mathcal{A}, \mathcal{R}, Ag, Src, T \rangle$ iff all of the following hold:

1. $\emptyset \in \mathcal{C}$ and every other node can be reached from \emptyset via \mathcal{V} ;
2. for any non-leaf node C , it is the case that $l(C) \in Ag$;
3. for any edge $v \in \mathcal{V}$, it is the case that $l(v)$ is of the form $c(a)$ where $a \in \mathcal{A}$ and $c \in \{+, -, s, h\}$;
4. for any edge $(C_1, C_2) \in \mathcal{V}$, if $l((C_1, C_2)) = c(a)$ for some $a \in \mathcal{A}$ and $c \in \{+, -, s, h\}$, then $Src(a) = l(C_1)$;
5. for all $e = (C_1, C_2) \in \mathcal{V}$, it is the case that $C_2 = C_1 \cup l(e)$;
6. for every leaf node C , it is the case that $l(C) = E$ where E is such that $cac\text{-}sub(\langle \mathcal{A}, \mathcal{R}, Ag, Src \rangle, \{ac\text{-}com^C\}) = \{(\mathcal{A}', \mathcal{R}', E)\}$.

The first constraint is that the starting point is where no decision has yet been made. All the cases considered can be reached from this starting point. The second constraint, together with the third constraint, forces the process to consider only the commitments made by one agent at a time. The fourth constraint represents the fact that a commitment on a certain argument can only be made by the agent who is its source. The fifth constraint represents carrying over commitments. The last constraint labels the nodes with the resulting collective extension. It computes the collective sub-framework semantics resulting from assigning the C-committed stable sub-framework semantics via individual acceptance functions for each agent.

Example 9 (Three agents, continued from Example 8). Consider the scenario described in Example 8—the case where *Acc* trusts *Prc*, and *Prc* trusts *Wit*, but *Wit* does not trust *Acc*. *Wit* does not need to condition his local framework, and therefore accepts a_3 and a_5 , and rejects a_4 . *Wit* may now decide whether to share or hide a_4 . If he hides a_4 , he only communicates the framework $\langle \{a_3, a_5\}, \emptyset \rangle$. Now, *Acc* considers the possibility that a_6 is acceptable and thus a_2 is not. *Acc* can then decide whether to share or hide a_2 , since he considers that the argument might not be accepted.

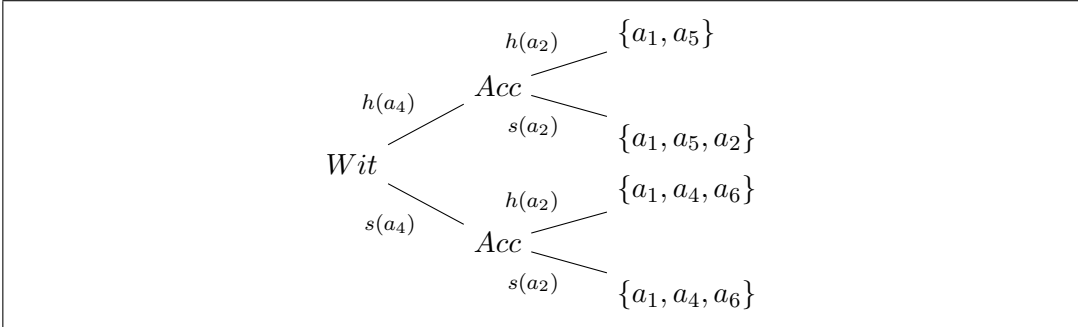


Figure 8: Commitment graph for the argumentation framework in Example 8.

Example 10 (Four arguments, continued from Example 6). *Recall running example 6. Assume that neither agent trusts the other. In Figure 9, we can see that in this case, the decisions of the agents have a major impact on the reasoning of the judge. Note also that in this case, the agents also have to decide which extension to accept before making any decisions about how to communicate this extension. Following the notation of Dauphin et al. [23], we represent this by using $+a$ when an agent commits to accepting an argument and $-a$ when the agent commits to rejecting it.*

However, in Figure 9, we notice that there is a lot of redundancy between some nodes. By applying a similar method to [23], we can reduce the graph to ensure that every node represents a meaningful commitment that isn't merely the logical consequence of a previous commitment.

We can define a reduction that collapses a node if for each child, the subgraphs that it can reach is isomorphic. We remove the node in question and replace it with one of its children, removing every other child from the original node. We first define what we mean by the subgraph that a given node generates.

Definition 12. *Let $(\mathcal{C}, \mathcal{V}, l)$ be a multi-agent commitment graph and $c \in \mathcal{C}$. We say that the subgraph generated by c is $(\mathcal{C}', \mathcal{V}', l')$ where \mathcal{C}' is the set of all nodes accessible with \mathcal{V} from c , where $\mathcal{V}' = \mathcal{V} \cap \mathcal{C}'^2$ and where l' is the restriction of l on \mathcal{C}' .*

Definition 13. *Let $(\mathcal{C}, \mathcal{V}, l)$ be a multi-agent commitment graph and $c \in \mathcal{C}$. We say that a node is reducible iff for every node $c' \in \mathcal{C}$ such that $(c, c') \in \mathcal{V}$, the subgraphs generated by the c' have an isomorphic relationship² to one other. In this case, we say that the graph where the subgraph generated by such a c has been replaced by the*

²Two graphs are isomorphic iff there exists a one-to-one mapping from one graph to the other graph that preserves the relation and labels.

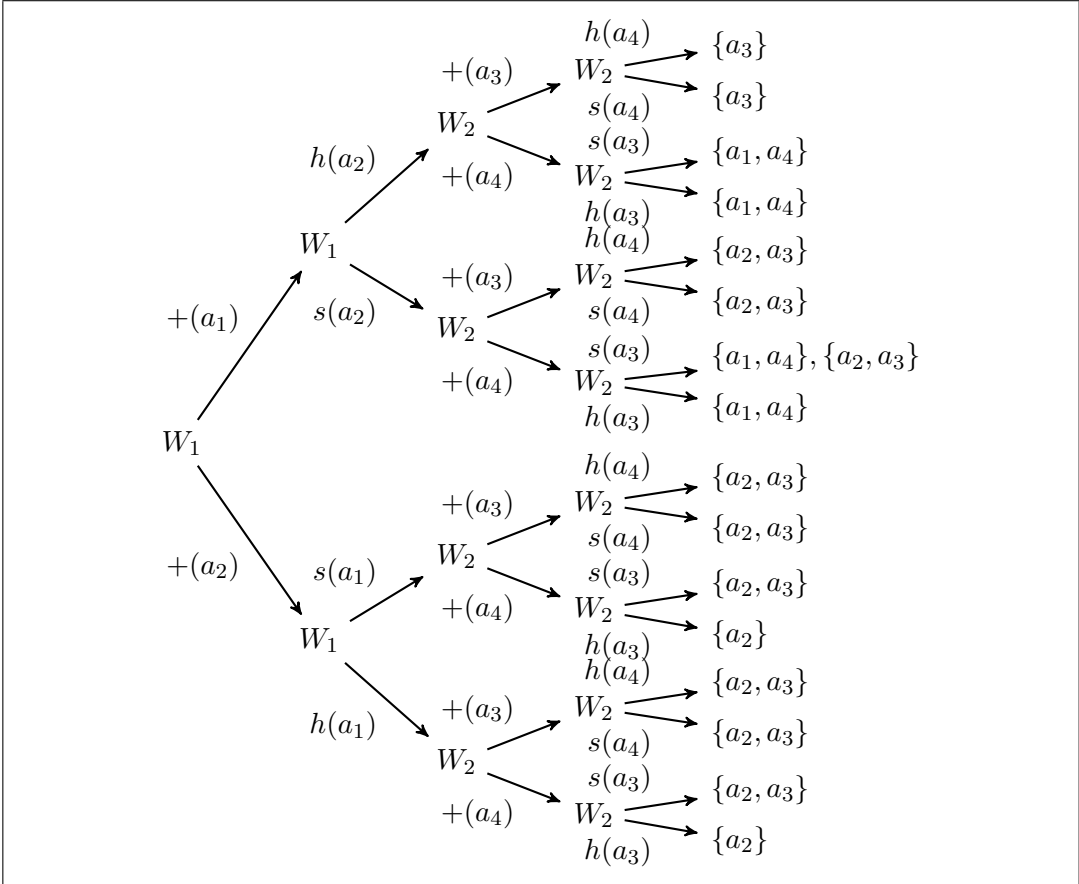


Figure 9: Multi-agent commitment graph for Example 10.

graph generated by the graph of its child c' is a reduction of the graph. If no such reduction exists, we say that the graph is minimal.

Example 11 (Four arguments, continued from Example 10). A minimal reduction of the decision graph from Figure 9 is depicted in Figure 10. Notice that every time a node used to be connected to two leaves labelled with the same extension, that node has now been replaced by a leaf labelled with that extension. Observe also that when W_1 committed to accepting a_2 , it did not matter whether W_1 hid or revealed a_1 . Therefore, the two subgraphs generated by those leaves could be merged, simplifying the resulting graph.

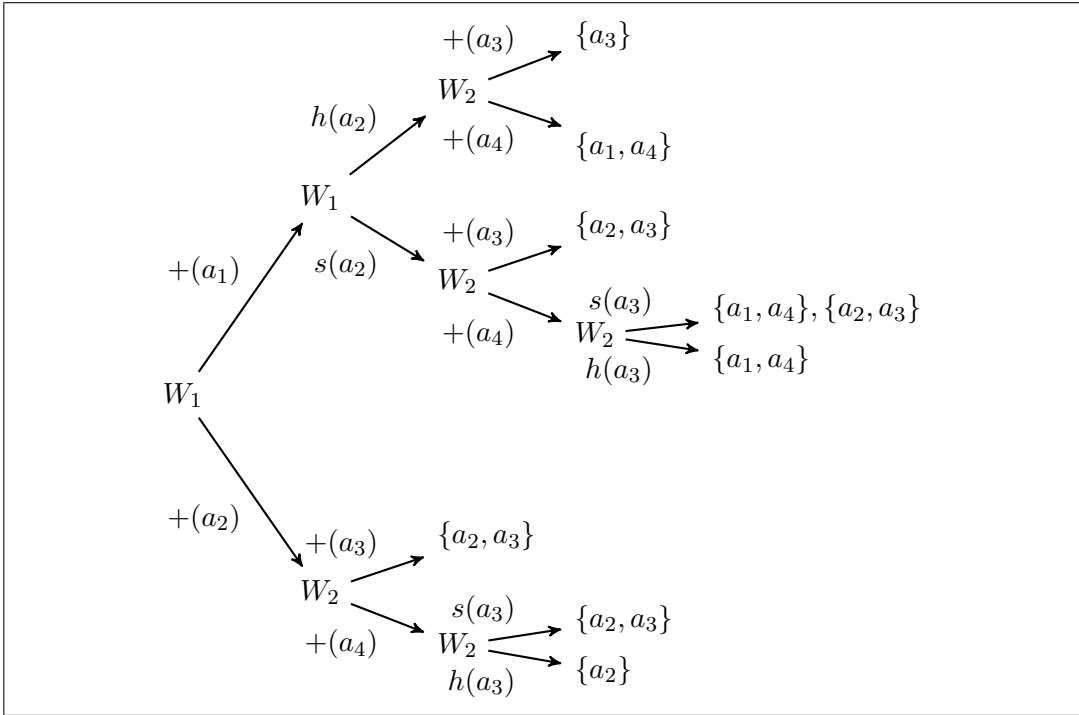


Figure 10: Reduction of the commitment graph in Figure 9.

6 Related and future work

In this section, we provide an overview of existing related research and some ideas for future work.

6.1 Abstract argumentation

The *Handbook* series on formal argumentation provides an up-to-date overview of the area. The first volume [11] presents the foundations of abstract and structured argumentation, and connects it to the rest of the argumentation literature. The second volume (to appear) offers extensions to abstract argumentation, analyses its dynamic aspects, and investigates the field at a meta level.

Dung [26] introduced the notion of admissibility-based semantics as a generalisation of stable sets. While the definitions in this article are based on the idea of stable sets, they can easily be adapted to other notions of acceptability, such as admissibility. Since the work of Dung, many other notions have been introduced, such as naïve-based semantics [16], weak admissibility [14] and strong admissibility [12].

It remains to be seen whether the approach described in this article generalises to all of these.

The idea of using sub-frameworks as an output of the semantics has been introduced previously in various contexts, most notably attack semantics [58] and argumentation frameworks with recursive attacks [10]. The interpretation and application described in this article are, however, novel.

Baroni *et al.* [13] had previously introduced the idea of local functions, and they have been used in subsequent work, most notably concerning principles [57], multi-sorted argumentation [47], and input/output argumentation [9]. In the work of Arisaka *et al.* [7], the sorts were interpreted as different agents and combined with sub-framework semantics. The work of Giacomin [27] and van der Torre *et al.* [56] have a similar interpretation for the sorts.

Dialogue semantics has been widely studied, and a comprehensive overview of the work carried out in this direction is provided by Caminada [19]. In these dialogues, two players exchange arguments from a given argumentation framework in order to prove or disprove the acceptability of a particular argument. These dialogues serve as a proof theory for the acceptability status (sceptical, credulous) of arguments with respect to various semantics, allowing one to prove the acceptability of an argument without requiring the computation of all the extensions. In this article, we considered a different kind of dialogue where all agents have their own sets of arguments and choose to hide or reveal them. We then observed how these choices affect the final verdict of an external observer. An interesting line of work would be to combine these two kinds of dialogues and provide a proof theory for acceptability statuses in the frameworks presented in this article.

Some researchers have followed a principle-based approach to the classification of argumentation semantics [12, 57]. A natural related first step towards the further development of multi-agent argumentation would be a definition of principles.

The work of Amgoud and Ben-Naim [1] aims to extract more information from argumentation frameworks by providing rankings pertaining to their degree of acceptability rather than sets of acceptable arguments. An interesting continuation of the work described in this article would be to investigate the effects of having agents share a ranking instead of a sub-framework. Then, the agents could also choose how much of their ranking to share, thereby hiding some of their arguments.

The dialogue semantics described in section 5 is based on the work of Dauphin *et al.* [23], which examines the decisions made while choosing an extension from a set of extensions. Another work of theirs [24] studies a similar structure for a different purpose. There, the focus is on detailing the process of getting from an argumentation framework to providing extensions to the semantics in order to allow different semantics to be combined in a meaningful way.

Many extensions of abstract argumentation frameworks have been proposed, together with more general notions of acceptance. Some example additions to the basic argumentation frameworks are preferences [2], support relations [21], abstract dialectical frameworks [18], and higher-order relations [10]. Abstract dialectical frameworks are discussed in more detail in the first volume of the *Handbook of Formal Argumentation* [17], while other extensions are discussed in the second volume of the *Handbook* (to appear).

While the agent-to-agent relation of trust is natural, we could alternatively define a function that maps an agent to a set of arguments it trusts [7], which adds to the expressiveness. Arisaka and Bistarelli [3] have studied the usage of a version of the agent-to-agent relation with an extra parameter for determining the mode of interaction for characterising dynamic collaboration among agents through defence delegation.

6.2 Merging argumentation frameworks and social argumentation

Whereas our model of multi-agent argumentation takes its inspiration from game theory, and it can be further developed towards coalitional game theory by introducing the arguments of sets of agents, an alternative approach to multi-agent argumentation takes its inspiration from voting theory, and more generally from social choice. One way to generalise our model towards such approaches is to consider a *Src relation* (rather than a function) between arguments and agents, such that two agents share the same argument. Likewise we can consider the sharing of attacks.

Sharing arguments:

Sharing an argument may mean that two arguments were put forward independently (so that there are two independent sources), or that one agent learned the argument from the other agent and copied it (so that there is one source and one copycat), or that two agents working together have prepared an argument using their combined knowledge (so that there is one source consisting of two agents).

Sharing attacks:

Attacks among arguments either do or do not depend on the agent. For example, in the former case, if two agents A and B both have arguments a and b , it may be the case that for agent A we have that a attacks b , and vice versa for agent B i.e. argument b attacks argument a . This could be interpreted using agent-specific preferences where agent A prefers a to b , whereas agent B prefers argument b to argument a .

We can relate this sharing of arguments and attacks also to the trust social network we introduced in Section 3. In an extreme case, it could be that if an

agent trusts another agent, then (s)he also incorporates his/her arguments and/or attacks into his/her own framework. Moreover, as we discuss in Section 3 of the article, if we assume the trust relation in Section 3 to be symmetric and transitive, then the equivalence classes of this trust relation may be called coalitions as well. If agents can share arguments, immediately the question arises: how is this related to coalitions? For example, do agents in a coalition share the same arguments? We believe this would be a natural assumption.

This can be related in different ways to existing theories of merging argumentation and social argumentation. For example, in social choice, the terms merging, fusion, voting and aggregation are often used interchangeably. Moreover, there are closely related approaches like negotiation with slightly different formal theories. There is a choice between combining the frameworks of the individual agents into a common framework by voting on the existence of arguments and attacks [22, 37], or making it so that they can agree on the framework and vote on the extensions [8, 20, 15, 55].

We believe that this raises the challenge of how to define a general research programme of multi-agent argumentation, of which all the above approaches are specific instances. Such a general theory would then explain precisely how a relatively simple procedure like voting can be compared to much more complex social phenomena like deliberation, negotiation, and argumentation.

6.3 Strategic dialogue games and 3+ multi-agent argumentation

In strategic dialogue games and agent persuasion (see [54] for a somewhat dated survey on the former, and [32] on the latter), multiple agents (most of the literature focuses on two parties) play a dialogue game, where the agents take turns in putting forward arguments with the aim of getting some arguments accepted (the goal of a proponent agent) or rejected (the goal of an opponent agent) in the end. While there has been some discussion of argumentation games under perfect information [44, 48, 43] that consider assigning payoffs to moves of putting forward a set of arguments and attacks, with the notion of equilibria then defined as in game theory, a dialogue game may not assume perfect knowledge of the environment on the part of the agents, thus opening up opportunities for opponent modelling, i.e. estimating their opponents' argumentation graphs, preferences, trusts, semantics and so on [4, 40, 52, 29, 30, 41] to gain strategic advantages.

Moreover, the arguments put forward may not always be truthful, since a greater strategic advantage may be gained by withholding information [45, 48, 29] or even by bluffing (common in Poker, Mafia/Werewolf, Mah-jong, and other imperfect information games). Strategic deceptive argumentation was first modelled explicitly in

[52]. A variation is found in [53] which relaxes the assumption of attack-omniscience (every agent knows argument-to-argument relations between two arguments precisely as long as the arguments are known to the agents). Other assumptions and certain anomalies in both studies were examined in [4], which also linked detected deception/honesty to changes in trustworthiness. The work of Kuipers and Denzinger [36] studied exploitation in logic-based argumentation that may arise from agents' differing logical inference capabilities. There is also a study on measuring the accuracy of lying/hiding detection based on observations on an agent's traits [35].

Multi-agent imperfect information argumentation among strictly more than two parties present other technicalities such as concurrency [6, 5] and collaboration [7, 3]. There has also been discussion of two-party perfect information dialogue games with an external observer, where the two agents aim to persuade the observer to accept some arguments by estimating the observer's belief about where argument-to-argument attacks are taking place [28].

In this article, we had a glimpse of opponent modelling, collaboration, information withholding, selective trust in other agents, and above all how these affect agent semantics, by letting the agent choose whether or not to take into account input argument(s) for computing its semantics, and what to share. As we observed in Example 6, Example 7 and Theorem 1, collective acceptance may not always match the stable semantics applied globally, depending on whom is trusted by which agents.

Some more recent work on persuasion also focuses on incorporating strategic aspects into natural language situations such as chatbots [33]. They combine domain modelling (predicting the arguments that could come up in the discussion), user modelling (representing the user's beliefs) and dialogue strategies (selecting the best argumentation moves for persuading the user). An interesting question is how to improve their system by incorporating the dialogue strategies outlined in this article.

7 Conclusion

Dung's abstract argumentation is the de facto standard for argumentation-as-inference. To bring it closer to the theories of argumentation-as-dialogue, we introduced agent interaction and dialogue into Dung's theory. This is a step towards a unified formal theory of argumentation covering both argumentation-as-inference and argumentation-as-dialogue, just like game theory is a unified theory for strategic and extensive games.

The starting point for multi-agent argumentation is the concept of conditional acceptance. In particular, an argument that an agent does not accept can be still be

put forward as part of the discussion. For example, agents can explain why they do not accept particular arguments by presenting counter-arguments to the unaccepted arguments, and they may even be willing to accept the argument if convinced by the other agents that their counter-arguments are wrong.

Multi-agent argumentation assigns arguments to agents, and associates individual acceptance functions with these agents. Multi-agent argumentation extends input/output argumentation in two ways. First, whether an agent accepts an argument put forward by another agent depends on the trust the agents have in one another, which is represented by a social network. Secondly, agents can decide whether or not to hide some of their arguments from the other agents. This concerns, in particular, the arguments they do not accept themselves.

An example from a court case illustrates the application of the above to legal reasoning. The agent abstract argumentation model distinguishes the global reasoning of judges from the local reasoning of the accused, prosecutors, witnesses, lawyers, experts and so on. The agents decide autonomously whether to trust the other agents in the sense that they take some of their arguments into account. Moreover, they decide autonomously whether to accept or reject their own arguments, and whether to bring their arguments forward in court. The arguments that are globally accepted by the judge are defined using a game theoretic equilibrium definition. The example distinguishes between various direct and indirect ways in which agents' arguments can be used against their other arguments.

A dialogue is a sequence of steps from the framework to the extensions where at each step of the sequence, agents can commit to accepting some arguments, or commit to hiding or revealing one of their rejected arguments. The revealed arguments are then aggregated and an external observer, in our example the judge, can compute which arguments are finally accepted at a global level.

The theory of multi-agent argumentation discussed in this article can be generalised in the way that is described in the *Handbook of Formal Argumentation*, for example by studying other semantics, defining principles, and by using structured argumentation theory, algorithms, and extensions that involve other concepts like preference and support. These further developments can be guided by our desire to bring the theory of abstract argumentation closer to existing theories of argumentation-as-dialogue. Several concrete proposals are discussed in the related work section of this article.

Acknowledgements

We thank Massimiliano Giacomin and our two anonymous reviewers for insightful feedback on an earlier version of this chapter. Leendert van der Torre acknowledges financial support from the Fonds National de la Recherche Luxembourg (INTER/-Mobility/19/13995684/DLAI/van der Torre)

References

- [1] Leila Amgoud and Jonathan Ben-Naim. Ranking-based semantics for argumentation frameworks. In *International Conference on Scalable Uncertainty Management*, pages 134–147. Springer, 2013.
- [2] Leila Amgoud and Srdjan Vesic. Rich preference-based argumentation frameworks. *International Journal of Approximate Reasoning*, 55(2):585–606, 2014.
- [3] Ryuta Arisaka and Stefano Bistarelli. Defence Outsourcing in Argumentation. In *Proceedings of the Seventh International Conference on Computational Models of Argument (COMMA)*, pages 353–360, 2018.
- [4] Ryuta Arisaka, Makoto Hagiwara, and Takayuki Ito. Deception/Honesty Detection and (Mis)trust Building in Manipulable Multi-Agent Argumentation: An insight. In *Proceedings of the Twenty-second International Conference on Principles and Practice of Multi-Agent Systems (PRIMA)*, pages 443–451, 2019.
- [5] Ryuta Arisaka and Takayuki Ito. Numerical Abstract Persuasion Argumentation for Expressing Concurrent Multi-Agent Negotiations. In *IJCAI Best of Workshops 2019*, 2019.
- [6] Ryuta Arisaka and Ken Satoh. Abstract Argumentation / Persuasion / Dynamics. In *Proceedings of the Twenty-First International Conference on Principles and Practice of Multi-Agent Systems (PRIMA)*, pages 331–343, 2018.
- [7] Ryuta Arisaka, Ken Satoh, and Leendert van der Torre. Anything you say may be used against you in a court of law: Abstract Agent Argumentation (Triple-A). In *Proceedings of the Ninth Workshop on Artificial Intelligence and the Complexity of Legal Reasoning (AICOL)*, pages 427–442, 2018.
- [8] Edmod Awad, Richard Booth, Fernando Tohmé, and Iyad Rahwan. Judgement Aggregation in Multi-Agent Argumentation. *Journal of Logic and Computation*, 27(1):227–259, 2017.
- [9] Pietro Baroni, Guido Boella, Federico Cerutti, Massimiliano Giacomin, Leendert W. N. van der Torre, and Serena Villata. On the input/output behavior of argumentation frameworks. *Artificial Intelligence*, 217:144–197, 2014.
- [10] Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Giovanni Guida. AFRA: argumentation framework with recursive attacks. *International Journal of Approximate Reasoning*, 52(1):19–37, 2011.

- [11] Pietro Baroni, Dov M Gabbay, Massimiliano Giacomin, and Leendert van der Torre. *Handbook of formal argumentation*. College Publications, 2018.
- [12] Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700, 2007.
- [13] Pietro Baroni, Massimiliano Giacomin, and Giovanni Guida. Scc-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168(1-2):162–210, 2005.
- [14] Ringo Baumann, Gerhard Brewka, and Markus Ulbricht. Revisiting the foundations of abstract argumentation–semantics based on weak admissibility and weak defense. In *Proceedings of the Thirty-Fourth International Conference on Artificial Intelligence (AAAI)*, 2020.
- [15] G. Bodanza and M. Auday. Social argument justification: some mechanisms and conditions for their coincidence. In *Proceedings of the Tenth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, pages 95–106, 2009.
- [16] Andrei Bondarenko, Phan Minh Dung, Robert A Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial intelligence*, 93(1-2):63–101, 1997.
- [17] Gerhard Brewka, Stefan Ellmauthaler, Hannes Strass, Johannes P. Wallner, and Stefan Woltran. Abstract dialectical frameworks. In Pietro Baroni, Dov M Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of formal argumentation*, volume 1, chapter 5, pages 237–285. College Publications, 2018.
- [18] Gerhard Brewka and Stefan Woltran. Abstract dialectical frameworks. In *Twelfth International Conference on the Principles of Knowledge Representation and Reasoning*, 2010.
- [19] Martin Caminada. Argumentation semantics as formal discussion. In Pietro Baroni, Dov M Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of formal argumentation*, volume 1, chapter 10, pages 487–518. College Publications, 2018.
- [20] Martin Caminada and G. Pigozzi. On judgement aggregation in abstract argumentation. *Autonomous Agents and Multi-Agent Systems*, 22(1):64–102, 2011.
- [21] Claudette Cayrol and Marie-Christine Lagasque-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 378–389. Springer, 2005.
- [22] Sylvie Coste-Marquis, Caroline Devred, and Sébastien Konieczny. On the Merging of Dung’s Argumentation Systems. *Artificial Intelligence*, 171(10-15):730–753, 2007.
- [23] Jérémie Dauphin, Marcos Cramer, and Leendert van der Torre. Abstract and concrete decision graphs for choosing extensions of argumentation frameworks. *Computational Models of Argument*, 2018.
- [24] Jérémie Dauphin, Marcos Cramer, and Leendert van der Torre. A dynamic approach for combining abstract argumentation semantics. In *Dynamics, Uncertainty and Rea-*

- soning, pages 21–43. Springer, 2019.
- [25] Alain Degenne and Michel Forsé. *Introducing social networks*. Sage, 1999.
 - [26] Phan M. Dung. On the Acceptability of Arguments and Its Fundamental Role in Non-monotonic Reasoning, Logic Programming, and n-Person Games. *Artificial Intelligence*, 77(2):321–357, 1995.
 - [27] Massimiliano Giacomin. Handling heterogeneous disagreements through abstract argumentation. In *International Conference on Principles and Practice of Multi-Agent Systems (PRIMA)*, pages 3–11. Springer, 2017.
 - [28] Davide Grossi and W. van der Hoek. Audience-based uncertainty in abstract argument games. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence (IJCAI)*, pages 143–149, 2013.
 - [29] Christos Hadjinikolis, Yiannis Siantos, Sanjay Modgil, Elizabeth Black, and Peter McBurney. Opponent modelling in persuasion dialogues. In *Proceedings of the Twenty-third International Joint Conference on Artificial Intelligence (IJCAI)*, pages 164–170, 2013.
 - [30] Emmanuel Hadoux, Aurélie Beynier, Nicolas Maudet, Paul Weng, and Anthony Hunter. Optimization of Probabilistic Argumentation with Markov Decision Models. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2004–2010, 2015.
 - [31] Catherine A Heaney and Barbara A Israel. Social networks and social support. *Health behavior and health education: Theory, research, and practice*, 4:189–210, 2008.
 - [32] Anthony Hunter. Towards a framework for computational persuasion with applications in behaviour change. *Argument & Computation*, 9(1):15–40, 2018.
 - [33] Anthony Hunter, Lisa Chalaguine, Tomasz Czernuszenko, Emmanuel Hadoux, and Sylwia Polberg. Towards computational persuasion via natural language argumentation dialogues. In *Joint German/Austrian Conference on Artificial Intelligence (Künstliche Intelligenz)*, pages 18–33. Springer, 2019.
 - [34] Trung Dong Huynh, Nicholas R Jennings, and Nigel R Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems (AAMAS)*, 13(2):119–154, 2006.
 - [35] Dionysios Kontarinis and Francesca Toni. Identifying Malicious Behaviour in Multi-party Bipolar Argumentation Behaviour. In *Proceedings of the Thirteenth European Conference on Multi-Agent Systems and Third International Conference on Agreement Technologies (EUMMAS/AT)*, pages 267–278, 2015.
 - [36] Andrew Kuipers and Jörg Denzinger. Pitfalls in Practical Open Multi Agent Argumentation Systems: Malicious Argumentation. In *Proceedings of the Third International Conference on Computational Models of Argument (COMMA)*, pages 323–334, 2010.
 - [37] Joao Leite and Joao Martins. Social abstract argumentation. In *Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, 2011.
 - [38] Beishui Liao. Toward incremental computation of argumentation semantics: A decomposition-based approach. *Annals of Mathematics and Artificial Intelligence*, 67(3-

- 4):319–358, 2013.
- [39] Lik Mui. *Computational models of trust and reputation: Agents, evolutionary games, and social networks*. PhD thesis, Massachusetts Institute of Technology, 2002.
 - [40] Nir Oren and Timothy J. Norman. Arguing Using Opponent Models. In *Proceedings of the Sixth International Workshop on Argumentation in Multi-Agent Systems (ArgMAS)*, pages 160–174, 2009.
 - [41] Alison R. Panisson, Simon Parsons, Peter McBurney, and Rafael H. Bordini. Choosing Appropriate Arguments from Trustworthy Sources. In *Proceedings of the Seventh International Conference on Computational Models of Argument (COMMA)*, pages 345–352, 2018.
 - [42] Henry Prakken. Historical overview of formal argumentation. In Pietro Baroni, Dov M Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of formal argumentation*, volume 1, chapter 2, pages 75–143. College Publications, 2018.
 - [43] A. Procaccia and J. Rosenschein. Extensive-form argumentation games. In *Proceedings of the Third European Workshop on Multi-Agent Systems (EUMAS)*, pages 312–322, 2005.
 - [44] I. Rahwan and K. Larson. Argumentation and game theory. In *Argumentation in Artificial Intelligence*, pages 321–339. Springer, 2009.
 - [45] Iyad Rahwan and K. Larson. Mechanism design for abstract argumentation. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1031–1038, 2008.
 - [46] Sarvapali D Ramchurn, Dong Huynh, and Nicholas R Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19(1):1–25, 2004.
 - [47] Tjitze Rienstra, Alan Perotti, Serena Villata, Dov M. Gabbay, and Leendert W. N. van der Torre. Multi-sorted argumentation. In *Theory and Applications of Formal Argumentation - First International Workshop, TFAA 2011. Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, pages 215–231, 2011.
 - [48] R. Riveret and Henry Prakken. Heuristics in argumentation: A game theory investigation. In *Proceedings of the Second International Conference on Computational Models of Argument (COMMA)*, pages 324–335, 2008.
 - [49] Jordi Sabater and Carles Sierra. Regret: reputation in gregarious societies. In *Proceedings of the Fifth International Conference on Autonomous agents and Multi-Agent Systems (AAMAS)*, pages 194–195, 2001.
 - [50] Jordi Sabater and Carles Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: Part 1*, pages 475–482, 2002.
 - [51] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial intelligence review*, 24(1):33–60, 2005.
 - [52] Chiaki Sakama. Dishonest Arguments in Debate Games. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA)*, pages 177–184, 2012.

- [53] Kazuko Takahashi and Shizuka Yokohama. On a Formal Treatment of Deception in Argumentative Dialogues. In *Proceedings of the Fourteenth European Conference on Multi-Agent Systems and Fourth International Conference on Agreement Technologies (EUMMAS/AT)*, pages 390–404, 2016.
- [54] Matthias Thimm. Strategic Argumentation in Multi-Agent Systems. *Künstliche Intelligenz*, 28(3):159–168, 2014.
- [55] Fernando Tohmé, G. Bodanza, and Guillermo R. Simari. Aggregation of attack relations: a social-choice theoretical analysis of defeasibility criteria. In *Proceedings of the Fifth International Symposium on Foundations of Information and Knowledge Systems (FoIKS)*, pages 8–23, 2008.
- [56] Leendert van der Torre, Tjitze Rienstra, and Dov Gabbay. Argumentation as exogenous coordination. In *It's All About Coordination*, pages 208–223. Springer, 2018.
- [57] Leendert van der Torre and Srdjan Vesic. The principle-based approach to abstract argumentation semantics. In Pietro Baroni, Dov M Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of formal argumentation*, volume 1, chapter 16, pages 797–837. College Publications, 2018.
- [58] Serena Villata, Guido Boella, and Leendert W. N. van der Torre. Attack semantics for abstract argumentation. In Toby Walsh, editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 406–413. IJCAI/AAAI, 2011.

THE LAW OF EVIDENCE AND LABELLED DEDUCTION: TEN YEARS LATER

DOV GABBAY

King's College, London, and University of Luxembourg
dov.gabbay@kcl.ac.uk

JOHN WOODS

University of British Columbia, Vancouver, Canada
john.woods@ubc.ca

Abstract

The purpose of this article is to reveal, through examples, the potential for collaboration between the theory of legal reasoning on the one hand, and some recently developed instruments of formal logic. Three zones of contact are highlighted.

1. The law of evidence, in the light of labelled deductive systems (LDSs), discussed through the example of the admissibility of hearsay evidence.
2. The give and take of legal debate in general, and regarding the acceptability of evidence in particular, represented using the abstract systems of argumentation developed in logic, notably the coloured graphs of Bench-Capon. This is considered through an imaginary example.
3. The use of Bayesian networks as tools for analysing the effects of uncertainty on the legal status of actions, illustrated via the same example

These three kinds of technique do not exclude each other. On the contrary, many cases of legal argument will need the combined resources of all three.

An earlier version of this paper was originally published in Φ News, Vol. 4, October 2003, pp. 5–46. A revised book version was published in D.M. Gabbay, Canivez, P., Rahman, S., and Thiercelin, A. (Eds.) *Approaches to Legal Rationality, Logic, Epistemology, and the Unity of Science 20*, 1st Edition., Springer 2010., pp 295–331.

1 Background: Logic and Law

In the first half of the past century, logic took a turn to the mathematical, and many were of the view that logic was the better for it. In the passage from Frege to Whitehead & Russell and on to the likes of Tarski and Church, first-order extensional logic would acquire the historically puzzling honorific “classical”, notwithstanding that Frege’s and W & R’s logics did not fill that bill.² One of the main nonclassical forces in this period lay in propositional extensions of classical logic for the modalities necessity and possibility. These were adapted in turn to the-so-called modalities of knowledge, belief, time and tense, and obligation and permission. Also important was the development of propositional logics in which the classical theorem mislabelled “*ex falso quodlibet*” is blocked.³ The theorem asserts that from a contradiction every sentence follows of necessity. It was often taken to mean that from a contradiction everything whatever can be *inferred*. It doesn’t mean that in fact, and it is not true. Although many a “paraconsistentist” logician was guilty of this confusion, there was great value in the attention they called to inference from inconsistent information. There is reason to believe that humans routinely reason from inconsistent background information without ever going completely to the dogs see Woods [117, 119]. Rational inconsistency-management remains an open problem in logic and computer science and yet an insufficiently recognized one [46, 47, 87, 88, 70]. It is a central problem in the logic of jury-trials and yet, there too, it wants for recognition and resolution [119, appendix G, “Inconsistently based verdicts”].

At the turn of that century, the universal Turing machine made its *début*, and there was launched an unending quest to build a machine that’s worth talking to. In time, it would be possible for computers to talk to computers of different kinds. What is sought now is a computer that can talk to us without having to be simulacra of us. Since the mere fact of human conversation is a standing invitation to voice differences of opinion, computer science has a large stake in analyzing human argument. This marks a return of the human individual to the focal exactions of formal methods [32, 33]. Modal logics acquired their quantificational wings and certain logics of deduction would, in the company of information theory, take the turn to pragmatics [32, 33, 71, 51]. Meanwhile the philosophy of mathematics sorted itself into the standard *collegia* of logicism, intuitionism and formalism. In what remained of the final half of the 20th century, logic would lose much of its historically acknowledged claim to be the one and only authoritative canonical framework for deductive thought. It is not that the very idea of it lost all credence, but rather that

²Frege’s was a second-order functional calculus harnessed to a theory of what we now call sets. Whitehead and Russell’s logic was a typed logic over propositional functions, hence not extensional.

³An accurate name is “*ex contradictione quodlibet*”.

no one theoretical claimant to the title managed to qualify itself for it. Not only had pluralism taken deep root in logic's deductive precincts [18], computer scientists had turned their attention to what mathematical logicians had long ignored. It was an immensely profitable turning in which, again, the reasoning agent was restored to formal consideration in the modelling of inference, decision-making, and tactical and strategic thinking. Arising from these freshly restimulated contexts were solid theories of nonmonotonic reasoning, defeasible reasoning, default judgement, autoepistemic reasoning, abductive logic and the sundry operations of AI [30]. Fruitful crossovers were wrought between the modal logics of belief, time and obligation, various of which skillfully negotiated and softened the older boundaries that had somewhat lazily dispersed mathematical philosophy to the camps of the classical, intuitionist and the formalist [74, 75].

Towards the end of the 1990s, logics of abductive inference started coming into flower [2, 76, 31, 81, 82, 84, 106, 53, 90, 89], and efforts would soon be made to model aspects of legal reasoning abductively [108, 115, 118]. The emphasis of much of this work fell on reasoning as a practical matter, and since the agents who reason practically are beings like us ensuing logics would in time take a naturalistic turn to the analysis of human reasoning. It would be a turn for the better for the logic of law. Abductive logic, too, would take an expressly naturalistic turn [86, 20] against an enlarging background of naturalistic approaches to inference more generally [116, 83, 85, 96, 29].

Meanwhile, a full-scale rebellion against the mathematicised formal logics of the day was launched by informal logicians who had taken up the task of restoring the systematic studying of fallacies to the research programmes of logic. If we date this resistance from the year in which Charles Hamblin's book *Fallacies* appeared [66], we would see soon after an emerging and prosperous interweaving of logic and epistemology, some of it very high levels of mathematical abstraction, and others more closely tethered to what happens on the ground of everyday thought and action [69]. In virtually all these often rivalrous iterations, there are unmistakable commonalities. In the main, the target of these myriad approaches was the human actor, making his way through life in real time as best he can with the resources at his command. This common orientation called for consideration of goals, actions, time and resources. A fruit of this widely shared focus was the cross-disciplinary readiness, even among rival theorists, to adopt from one another aspects of their proceedings in hopes of using them to greater advantage in their own respective approaches [57]. Researchers would often approach problems in their own respective domains of enquiry by modelling them on the way theories in other domains treated

the problems that cropped up there.⁴ One day in London, Ray Reiter remarked to the present authors, “It is deliciously wild! Everyone is eating everyone else’s lunch!” He did not say this complainingly.⁵

An earlier and also important influence on the development of informal logic was the appearance in 1958 of Stephen Toulmin’s *The Uses of Argument* which, among other things, offered a more complex representation of argument structure than the more standard premiss + premiss + conclusion model such as would be found in such textbooks as Copi [23]. Known as the Toulmin Model, it picked out features of legal reasoning which Toulmin believed to be generalizable to arguments of all subject matters [101]. Although the Toulmin Model has continued to play well in theories of argument, perhaps a more substantial jolt to received opinion was delivered in Toulmin’s primer on the philosophy of science in the Hutchison Library series in 1953. In it Toulmin admonished theorists of inductive and probabilistic reasoning for their over-use of Bayesian methods, especially in contexts for which they are especially ill-suited. *The Uses of Argument* roiled mid-century thought about argument, to such an extent as to have brought it about that Stephen Toulmin was analytic philosophy’s “most refuted author”.⁶

Nineteen ninety-two marks a significant step in the logical investigation of legal reasoning, with the establishment of the journal *Artificial Intelligence and Law*, currently edited by Kevin Ashley, Trevor Bench-Capon and Giovanni Sartor. Soon after, T. F. Gordon’s monograph on AI modelling of procedural justice would appear [63]. In 2001, the present authors announced a research program in what came to be known as the practical logic of cognitive systems.⁷ While not explicitly focused on either AI or the law, it was so structured as to be amenable to such uses. The turn was taken in 2003 in the first iteration of the present study in Φ *News* 4, 5–46. While it bore the same main title as does the present version, it carried a subtitle which no longer applies — “A Position Paper”. That same year there appeared the first volume of our omnibus work, *A Practical Logic of Cognitive Systems* under the

⁴For example, the so-called Woods-Walton Approach to fallacy theory found profitable assistance in intuitionist logic, graph theory, relatedness logic, aggregate theory, plausibility logics, dialectics, dialogue logics, and decision theory.

⁵Also significant is the relaxation of logic’s mathematical preclusion of the physical advocated in Putnam [95]. The birth of quantum logic would have irritated Frege, but Putnam’s suggestion that logic should be seen as a natural science would have infuriated him. But when it takes the practical turn to objects of nature, this is precisely what the logic of inference turns out to be. In due course, quantum logics would be a flourishing enterprise [25, 15, 28].

⁶In the words of William Alston in 1962, when introducing Toulmin to a packed house at the University of Michigan. We find it surprising that Toulmin’s probabilistic deviations didn’t raise much dust in the philosophy of science.

⁷Gabbay DM and Woods J. 2001. The new logic, *Logic Journal of the IGPL*, 9, 157–186.

title, *Agenda Relevance: A Study in Formal Pragmatics* (2003b), in two sections of which we took up the question of legal relevance (5.3 and 9.7). Two years later legal relevance and legal presumption were taken up in volume 2. *The Reach of Abduction: Insight and Trial* (2005) in sections 8.4 and 8.5.⁸ A significant step in the modelling of legal reasoning using AI techniques was Walton [107]. Building on earlier work on the dialogical structure of legal reasoning [104, 105] more AI-oriented work would follow [107, 108, 109, 110], along with co-authored work [64, 97, 111, 109]. Further work in this area includes Gordon [63], Verheij [103], Keppens [73] and Prakken and Sartor [93]. A well-received reference work on this subject is Bongiovanni, Postema, Rotolo, Sartor and Walton, *Handbook of Legal Reasoning and Argumentation* [14].

Once logic has evolved in this direction and has developed new logical tools for this purpose, these same kind of new logics and new tools can usefully be adapted to the consideration of similar issues in the law.

Here lies the connection between logic and law. We can say without serious exaggeration that the interface of logic and law is going to be central to the further advancement of logic in the next twenty years. If only we can bring the respective communities together and make them aware of their potential! This is the purpose of this article.

We envisage the following main benefits to the law community, in addition to the benefits from existing logical tools and aids available from Artificial Intelligence.

- The proper LDS logic tailored for law of evidence and other judicial arguments can help articulate and clarify (hidden) intuitive common sense principles behind existing practices.
- The LDS methodology includes a system of labelling and stylised hierarchical movements which have logical content. This kind of hierarchy can be added to legal specification formats thus giving a better specification language for law without sacrificing the use of ambiguities and variety of interpretations.

It is astonishing to realize that very few people are aware of the true potential of the interaction of the new logics and law. There are many reasons for that, most of them social. The new developments in logic are slow to spread around even among

⁸Shortly after, our publisher adopted a new business model and changed course markedly, and among other things, shut down the venerable “Yellow Series”, Studies in Logic and the Foundations of Mathematics. They denied us our continued use of the name “A Practical Logic of Cognitive Systems”, so a third volume by Woods — *Errors of Reasoning: Naturalizing the Logic of Inference* — would appear in 2013 with College Publications under the new series title “Logic and Cognitive Systems”.

logicians, and certainly among researchers in legal reasoning and legal theory, many of whom still think of “logic” as “Aristotelian syllogism”.⁹

Some bridging work between law and logic has been done by C.H. Perelman [91], who kept in touch with both logicians and judges and lawyers, arguing that logic should play a different — more restricted — role. But when Perelman wrote, the new logical tools were not as available as they are now; and such as were available, Perelman made no use of.

The rise of Horn clause logic programming in the 1980s has helped turn some logicians in the direction of the law, but early attempts to apply logic to law, such as the formalisation of the British Nationality Act [99], has rightly drawn a strong critical reaction from the legal community on the ground that Horn clause logic is not rich enough to allow for the wealth of nuances and interpretations/explanation/revision so common in legal reasoning. See also [1] by Judge Ruggero J. Aldisert.

This criticism may have been valid in 1980, the objection is no longer valid now, especially in view of many advances made in logics of practical reasoning and argumentation.

Logic programmers and deontic logicians have had a somewhat earlier interest in law, have their own conferences and journals [27]. But we doubt if they are aware as a community of all relevant developments in logic. They appear not to realize (or believe) that law is an area of potentially evolutionary significance to logic.

Still valuable are survey works by two key researchers in the area, Trevor Bench-Capon’s [11] survey article for the *Encyclopaedia of Computer Science and Technology* and Henry Prakken’s book [92], *Logical Tools for Modelling Legal Argument*.

⁹It is instructive to read the following passage on legal reasoning from the July 2003 edition of a basic textbook on legal philosophy, widely taught in the UK (J. W. Harris, *Legal Philosophies*, p 213):

“It is far from easy to get a comprehensive view of the subject [of legal reasoning]. Most writers who have discussed legal reasoning have either concentrated on the form as distinct from the substance of justificatory arguments, or else dealt with only part of the subject. Two forms of argument, the deductive and the inductive, have generally been considered inapposite characterizations of legal argument. Some take the view that deductive argument – from major and minor premises to a logically necessary conclusion – is inappropriate even in clear cases. This may be asserted on the general ground that deductive arguments only hold true of factual propositions not of norms; or on the more specific ground that even the clearest rule may be held not to apply to a case where that would frustrate the purpose of the law or produce absurd consequences, and the decision whether this so or not cannot be dictated by logic. On the other hand, reasoning in clear cases seems very close to deductive reasoning – here is a speed-limit rule applying to all car drivers, I am a car driver, so it applies to me. Even in unclear cases, it can be contended that the form of the argument is deductive, since what is at issue is which of competing rulings should be adopted, granted that the winner will be applied deductively in all cases of the present type – although here our major concern will be with the substantive arguments which dictate choice among the rulings.

Prakken's book, especially, takes note of many of the new developments in logic, and argues very strongly in favour of the theoretical connectedness of logic and law. He especially highlights the new developments in defeasible and non-monotonic logics and reasoning from inconsistent data. However, he is unaware of the methodology of labelled deductive systems which subsumes the logic of legal reasoning, among many others, as a special case. More importantly, Prakken believes that 'logic should be regarded as a tool rather than as a model of reasoning', [92, Section 1.4]. Furthermore, the entire approach to date of the community to logic and law is further restricted by the view that:

To understand the scope of the present investigations it is important to be aware of the fact that the information with which a knowledge-based system reasons, as well as the description of the problem, is the result of many activities which escape a formal treatment, but which are essential elements of what is called 'legal reasoning'. In sum, the only aspects of legal reasoning which can be formalised are those aspects which concern the following problem: *given* a particular interpretation of a body of information, and *given* a particular description of some legal problem, what are then the general rational patterns of reasoning with which a solution to the problem can be obtained? With respect to this question one remark should be made: I do not require that these general patterns are deductive; the only requirement is that they should be formally definable. [92, p. 6]

Thus modelling the legal theory of evidence (which decides what 'body of information' we are 'given') still remains beyond the horizon of some current research in logic and law. In what follows, on the contrary, we shall develop a case study that will show just how important this area is.

A recent key collection of papers by Marylin MacCrimmon and Peter Tillers [80] indicates very lively activity in law and logic. However, most of the papers take a fuzzy logic, uncertainty and probabilistic approach (in the sense of [100, 65]). See also [94] and the references there.

We must here add that the Bayesian reasoning community is actively involved in (Bayesian) logic and law. This is because of several high visibility court cases and evidence where probabilities are used. Part of the problem is that the probabilistic reasoning community is not so interactive with the ordinary logic communities (and so we also need to bring logic and probability together as part of our own ongoing work). However, for reservations about the reliability in their present formulations of the Bayesian norms for legal reasoning, see [118]. Suitably interpreted the theory

of Labelled Deductive Systems is fully compatible with probabilistic reasoning and networks.

In the sections to come we examine some case studies to show how the new logics can play a role in the area of evidence and legal reasoning.

2 Legal Theory of Evidence and the New Logics

Our purpose here is to show how the new labelled logics, arising from research in computer science, can be applied to the legal theory of evidence. For a sample of Labelled Deductive Systems, see [37]. For the original monograph, see [34].

2.1 Some Labelled Logic

We start with logic. One of the most well known resource logics is linear logic [62]. In this logic, the databases are multisets of formulas and each item of data must be used *exactly once*. So, for example, we have

$$A, A \rightarrow B \vdash B$$

But

$$A, A \rightarrow (A \rightarrow B) \not\vdash B$$

This is because two copies of A are needed here, and we have only one. The proof would run as follows:

1. $A \rightarrow (A \rightarrow B)$, assumption
2. A , assumption
3. $A \rightarrow B$, from 1 and 2 using the rule of modus ponens.
4. B , from 1 and 3, using the rule of modus ponens.

In this proof, 2. is used twice.

To make this example more concrete, let

- A = having a drunken driving conviction
- B = driving licence suspended.

Then $A \rightarrow (A \rightarrow B)$ means that two convictions entail suspension (and of course you cannot count the same conviction twice!).

Linear logic allows for the connective $!A$, which means that A can be used as many times as needed.

Thus

$$!A, A \rightarrow (A \rightarrow B) \vdash B.$$

Let us modify the logic a bit¹⁰ and add the connective $\heartsuit A$: $\heartsuit A$ means that we can use A if we ask and get permission from some meta-level authority. So we can write

$$\heartsuit A, A \rightarrow (A \rightarrow B), \text{ permission given} \vdash B.$$

There is a mixing here of object level and meta-level features. Such logics are best expressed as labeled deductive systems (LDS) [34, 37]. A labelled system is comprised of formulas and labels. The labels contain additional information relating to the formulas. For example an item of data (called a *declarative unit*) may have the form

$$\Delta : \text{John has cancer.}$$

Δ can be a medical file with data confirming the fact that John has cancer. This fact can be used in certain situations of legal argument; e.g. to attempt to release John from prison. The reasoning governing Δ is medical, while the reasoning governing the release from prison is legal. Labelled logic is the methodology of how to use such mixed reasoning.

We have in LDS the following form of modus ponens:

$$\frac{t : X, s : X \rightarrow Y, \varphi(s, t)}{f(s, t) : Y}$$

Here t, s are labels (their nature and mode of handling are defined in the system), which can be themselves entire databases; φ is meta-predicate indicating that there is the permission to apply modus ponens (φ is called the compatibility predicate); and f is a function giving the new label of the result Y .

Going back to our example, we write

1. $s : (A \rightarrow (A \rightarrow B))$, where s represents here a body of legal background data on how the substantive law of
“two drunken driving convictions \rightarrow licence suspended”
has been established.

¹⁰See footnote 30 for an anagram example.

2. $t : A$, where t is a file indicating the data establishing the facts of the drunken driving incident.
3. $\varphi(s, t)$ is a meta-level argument looking into s and t and arguing that, although we have here only *one* incident of drunken driving, the intention of law (see file s) and the severe circumstances of the incident (see file t) call for suspension (that is, permission to count as two incidents is granted).
4. $f(s, t) : A \rightarrow B$, by modus ponens from (1), (2), (3).
 $f(s, t)$ is a file containing the arguments present in granting permission, i.e.,
 $f(s, t) = t + s + \varphi$
5. $f(s, t) : B$, by modus ponens from (4) and (5).
So formally, we have $f(f(s, t), t) = f(s, t)$.

We now show a further connection with the law of evidence.

One important feature of LDS is that it regulates the admissibility of data into the database together with the label it is permitted to have. In fact, using φ we can diplomatically admit a datum D into the database with a label “don’t touch”, with the effect that φ will never give permission to use it.

These kinds of logics were developed to accommodate the needs of artificial intelligence and the logic of language. It is surprising how well these logics fit the needs of theories of evidence.

Imagine a database (Barclays Bank) containing data about a customer. One kind of data includes home telephone number, mobile telephone number, etc. Assume that a security protocol will allow only certain individuals at the Bank to enter such data and it is up to them to decide whether to ‘admit’ an additional number. Suppose I call Barclays bank, identify myself and ask the representative to add my mobile number to the database. The representative will ask me some questions (usually mother’s maiden name). If correct answers are given, he will add (admit) the additional telephone number. If he is still uncomfortable with my identity (for whatever reasons) he can refuse to do so. We doubt, however, that he has the authority to decide to accept the phone number even if we fail to answer the questions correctly. In other words, security protocols allow the representative to refuse admissible data but do not allow him to overrule and accept non-admissible data!

2.2 What Some Books on Evidence Say

Let us go now to the website and to the book of Steve Uglow.

In his web course notes, and presumably also in his book, he says:

“Evidence is about regulating the information produced at a trial.

- What are the general principles regarding this?
- What are exclusionary rules?
- What logical processes are involved?”

In our labelled logic we can phrase these points as

- With what label do we insert the new data (evidence) in our database?

The challenge of this area to the research community is made clear at the very first paragraph of Uglow’s 725-page book on evidence [102] (*Textbook on Evidence*, 1997)

“The law relating to evidence is a strange and unruly beast. It is unruly because, first, it refuses to fit into any easy structure for analysis and exposition and, second, it often adopts the characteristics of an uncharged minefield, by which is meant that any set of facts has the potential of throwing up evidential problems, not just of one but of several types, often unforeseen. It is strange because it fulfils different functions than the familiar areas of substantive law. It is in such areas that we see legal rules at their most visible, dealing with the *consequences* of facts – if a contract is broken, damages are paid; if a theft is committed, punishment is imposed. Damages, imprisonment and other civil and criminal remedies are the sanctions accompanying rules which require or prohibit certain types of conduct or which lay down conditions under which that conduct can take place. These rules are often referred to as the substantive law. Within most contested trials, such rules form the background to the case but play little part since there is no conflict over the substance of the rule. We know what the rule says and what the consequences of a breach will be: if there has been a road accident and a driver has been negligent, damages for personal injuries will be paid to any plaintiff; if a sane defendant intentionally kills another person, he or she will be prosecuted and generally receive a life sentence.

But the real conflict in a court, before any substantive rule is brought to bear, is about establishing the facts: was the driver negligent? Did the defendant cause the victim’s death? What happened? The law of evidence is not about determining the consequences of facts but about establishing those facts. In a contested trial, under the common law system of justice, the opposing parties will present differing, sometimes

diametrically opposed, views of the same event. Having listened to these accounts, the trier of fact must decide what the facts are. It is this problem as to how ‘facts’ are established with which the law of evidence is concerned: what information can be presented to the court’ through what means; how does a court decide whether that information proves whether an event happened in a particular way or not? Such rules, alongside the rules of civil and criminal procedure, can be described, not as *substantive*, but as *adjectival law*.¹¹

This means that these rules attach themselves to and qualify the operation of a substantive rule but never, by themselves, directly decide the rights and wrongs of any issue. The law of evidence qualifies the operation of a substantive rule because it controls the flow and nature of the information which can be presented to the court. Indirectly, of course, the law of evidence can be decisive since the outcome of a case can depend on whether a particular item of evidence is allowed to be presented to the court or not. For example, a guilty verdict or an acquittal can hang on whether the prosecution can meet the preconditions for the admissibility of a confession in a criminal trial; in a civil case where the

¹¹This is our footnote.

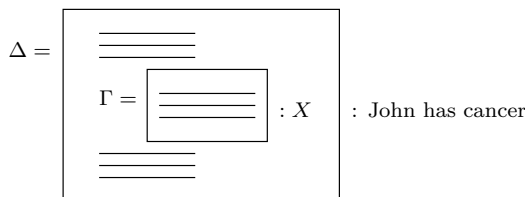
Note that a substantive law in labelled logic looks like $s : A \rightarrow (A \rightarrow B)$. Facts look like $t : A$. We can also have other testimony allowing for $t' : \neg A$. The rule that decides in *LDS*, whether to deduce A or $\neg A$ given say, $t_1 : A, t_2 : A, t_3 : \neg A$ is called a *flattening rule*. More precisely, a flattening rule tells us, given $t_i : A$ and $s_j : \neg A$, what is the resultant labels $t : A$ and $s : \neg A$. So, for example, if t_i, s_j are reliability measures of various sources supporting A and $\neg A$ respectively, t and s might be some averages.

What Professor Uglov calls here *Adjectival Law*, means in *LDS* the logic for reasoning *inside* the label t . For example, t may contain medical evidence and a lawyer may attack that!

If we take our example

Δ : John has cancer,

Δ may be a medical file about John. Δ may contain among other things an expert opinion of a certain Dr. Smith, giving a statement $\Gamma : X$, there X is the Doctor’s statement and Γ is another file showing Dr.Smith is a world expert on this kind of cancer. A lawyer wishing to attack Δ might choose to attack Γ (i.e., Dr. Smith’s credentials are false), thus weakening the value of X and overall weakening Δ . So we have a structure like



weight of the evidence is evenly balanced, the decision may hinge on the question as to where the burden of proof rests.

Many of these evidential issues seem very technical to a layperson and, especially in criminal trials, to exclude relevant and important information from the proceedings. Examples might be given of the rule against the admission of hearsay evidence — a witness would be usually prevented from testifying that the victim, now dead, had identified the accused as the assailant; similarly the jury would rarely be allowed to hear about any previous convictions of the defendant. But these are not technicalities for their own sake and reflect the nature and characteristics of the common law trial.”

Put in the language our own LDS, what Uglow is saying in this passage is that: Given the situation

$$t : A, s : A \rightarrow B$$

he calls $A \rightarrow B$ ”substantive law”, (in logic it is called a “rule” or a “ticket”), and calls A the facts (called minor premises in logic), then the main part of the theory of evidence is whether to admit A into the database (i.e., establish A as a fact) and with what label t ? t may be a label supporting A and what the book calls “adjectival law” is the theory (logic) of evidence.

Here now is another basic textbook on evidence [26], I. H. Dennis, *Law of Evidence*, 1999.¹² He says (pages 4–6)

B. Concepts and Terminology

The law of evidence uses a number of concepts which are fundamental to an understanding of the subject. This section attempts to introduce these concepts by stating a number of general propositions about them and about their relationships. The propositions are stated in summary form, with more detailed explanation given later.¹³

1. Evidence must be *relevant* in order for a court to receive it. This means that it must relate to some fact which is a proper object of proof

¹²Dennis also says in his introduction “Evidence is a notoriously difficult subject to organize in any logical basis”.

¹³Also important from a dialogical perspective are works by Douglas Walton on evidence in law. See Walton [105, 108, 110]. Woods emphasizes the artificialities of legal evidence in Woods [118, chapters 8 and 10].

in the proceedings.¹⁴ The evidence must relate to the fact to be proved in the sense that it tends to make the existence (or non-existence) of the fact more probable, or less probable, than it would be without the evidence. A simple example is a case where a fact to be proved is the identity of the accused as the person who stole certain goods. Evidence that the goods were found in the accused's house is relevant because it makes the existence of the fact that he is the thief more probable.

2. Evidence must also be *admissible*, meaning that it can properly be received by a court as a matter of law. The most important rule of admissibility is that the evidence must be relevant; irrelevant evidence is always inadmissible. Generally speaking evidence that is relevant is also admissible, but certain rules of law prohibit the reception of certain types of evidence, even though the evidence is relevant. An example is the rule against hearsay evidence, which, broadly speaking, forbids the reception of evidence of a statement made by a person on another occasion when the purpose of adducing¹⁵ the evidence is to ask the court to accept that the statement was true. These rules are often called the *exclusionary rules*, to indicate their function of excluding certain evidence from the court's consideration. The rules are complex because they are often accompanied by exceptions, some of which may be narrow and precisely defined, others may be in broad and flexible terms.

3. In criminal cases, in addition to exclusionary rules, there is also *exclusionary discretion*. A trial judge may exclude prosecution evidence that is relevant and admissible (in the sense that it is not excluded by an exclusionary rule) in the exercise of a discretion conferred on him by the common law or by section 78 of the Police and Criminal Evidence Act 1984 (PACE). The statutory discretion is to prevent the admission of the evidence from adversely affecting the fairness of the proceedings. The main application of the common law discretion is to exclude evidence the prejudicial effect of which outweighs its probative value. Probative value refers to the potential weight of the evidence (see next paragraph), whereas prejudicial effect refers to the tendency of evidence to prejudice the court against the accused, so as to lead the court to make findings

¹⁴The facts which are proper objects of proof are sometimes called material facts, but materiality is a slippery term which can be used with more than one meaning. See the discussion in the text below.

¹⁵"Adducing" evidence is a term often used to denote the process of presenting evidence to a court in one of the approved forms, most commonly in the form of the testimony of a witness.

of fact against him for reasons not related to the true probative value of the evidence.¹⁶

4. At the end of a contested trial the court will have to evaluate the relevant and admissible evidence that it received. The *weight* of the evidence is the strength of the tendency of the evidence to prove the fact or facts that it was adduced to prove. This is a matter for the tribunal of fact to decide. In civil cases the judge who tries the case is generally the judge of issues of both law and fact. In criminal cases the tribunal of fact is different according to whether the case is tried on indictment or summarily. The jury is the tribunal of fact for cases tried on indictment. In summary trial the magistrates (justices) deal with issues of both law and fact; lay magistrates have the guidance of their clerk on questions of law. This book uses the term “factfinder” to refer generally to a tribunal of fact, unless the context requires a specific reference to a judge, jury or magistrate. When a factfinder has to determine the weight of evidence it will examine carefully, amongst other things, the *credibility* and *reliability* of the evidence. These terms are not always used with a consistent meaning. Credibility is most commonly used in connection with the testimony of a witness and refers to the extent to which the witness can be accepted as giving truthful evidence in the sense of honest or sincere testimony. Reliability refers most commonly to the truthfulness of testimony in the sense of its accuracy. Honest witnesses may sometimes give evidence that is inaccurate; mistaken evidence of identification by eyewitnesses is a classic example.¹⁷

Note here the central role played by the notion of *relevance*. This is also an AI and natural language concept. It is no accident that the first book of our series of books on cognitive systems is a book on relevance [50].

3 Case Study: Hearsay Case, *Myers v DPP*

We begin by quoting from [3, p. 133].

A good statement of the hearsay rule was given originally in *Cross on Evidence*, [24].

¹⁶In other situations a judge can find clearly probative evidence to be irrelevant in law if it would be too difficult for the jury to understand or would take too long for the jury to hear. (“Justice delayed is justice denied.”) Woods [118, pp. 182–283].

¹⁷For some of the difficulties with eyewitness testimony, see Loftus [77] and Loftus *et al.*, [78].

“An assertion other than one made by a person while giving oral evidence in the proceedings is inadmissible as evidence of any fact asserted”.

Allen continued on page 135:

“Hearsay law has been described as ‘exceptionally complex and difficult to interpret’ [98]. What we need is a method of approach to the subject which will enable us to understand why some cases were decided as they were and why others are open to criticism. Above all, we need a technique [our comment: i.e., logic] for thinking about hearsay, . . .’.

We now examine a key case, which seems to be quoted in every textbook on Evidence (and hearsay). This is a case of *written statements*, which may fall under hearsay law.

We quote two descriptions of this case, one from [72] and one from [102], and then we model the arguments as quoted in [102].

We begin with [72, pp. 250–252]

(b) Written statements

The leading case on written hearsay is *Myers v DPP* ([1965] AC 1001). The appellant was convicted of offences relating to the theft of motor cars. He would buy a wrecked car, steal a car resembling it, disguise the stolen car so that it corresponded with the particulars of the wrecked car as noted in its log book, and then sell the stolen car with the log book of the wrecked one. The prosecution case involved proving that the disguised cars were stolen by reference to the cylinder-block numbers indelibly stamped on their engines. In the case of some cars, therefore, they sought to adduce evidence derived from records kept by a motor manufacturer. An officer in charge of these records was called to produce microfilms which were prepared from cards filled in by workmen on the assembly line and which contained the cylinder-block numbers of the cars manufactured. The Court of Criminal Appeal held that the trial judge had properly allowed the evidence to be admitted because of the circumstances in which the record was maintained and the inherent probability that it was correct rather than incorrect. The House of Lords held that the records constituted inadmissible hearsay evidence. The entries on the cards and contained in the microfilms were out-of-court assertions by unidentifiable workmen that certain cars bore certain cylinder-block numbers. The officer called could not prove that the records were correct and that the numbers they contained were in fact the numbers on the

cars in question. Their Lordships, however, were divided as to whether the evidence should be admitted by the creation of a new exception to the hearsay rule.¹⁸ Lords Pearce and Donovan were in favour of such a course, but the majority, comprising Lords Reid, Morris and Hodson, declined to do so, being of the opinion that it was for the legislature and not the judiciary to add to the classes of admissible hearsay.¹⁹ It was argued before the House that the trial judge has a discretion to admit a record in a particular case if satisfied that it is trustworthy and that justice requires its admission. Lord Reid, while acknowledging that the hearsay rule was ‘absurdly technical’, held that ‘no matter how cogent particular evidence may seem to be, unless it comes within a class which is admissible, it is excluded . . .’

The actual decision in *Myers v DPP* was reversed by the Criminal Evidence Act 1965, which provided for the admissibility of certain hearsay statements contained in trade or business records. Although the 1965 Act was repealed by the Police and Criminal Evidence Act 1984, ss 23 and 24 of the Criminal Justice Act 1988 are wider in scope than the provisions of the 1965 Act and provide for the admissibility of first-hand hearsay statements in documents generally as well as hearsay statements contained in documents created or received by a person in the course of, inter alia, a trade or business. The principles enunciated in *Myers v DPP*, however, remain of importance in relation to hearsay statements falling outside the statutory exceptions. Over 25 years later, another majority of the House of Lords, in *R v Kearley*,²⁰ although of the opinion that there may be a case for a general relaxation of the hearsay rule, affirmed the majority view in *Myers v DPP* that the only satisfactory solution is legislation following on a wide survey of the whole field.

*Patel v Comptroller of Customs*²¹ also illustrates the application of the hearsay rule to written statements. The appellant was convicted of mak-

¹⁸The Lords were unanimous in dismissing the appeal on the grounds that the other evidence of guilt being overwhelming, there had been no substantial miscarriage of justice.

¹⁹The minority view, that it was within the provenance of the judiciary to restate the exceptions to the hearsay rule, was adopted by the Supreme Court of Canada in *Ares v Venner* [1970] SCR 608. See also per Lord Griffiths in *R v Kearley* [1992] 2 All ER 345, HL at 348.

²⁰[1992] 2 All ER 345, HL, per Lords Bridge, Ackner and Oliver at 360–361, 366 and 382–383 respectively.

²¹[1966] AC 356, PC. See also *R v Sealby* [1965] 1 All ER 701 and *R v Brown* [1991] Crim LR835, CA (evidence of a name on an appliance inadmissible to establish its ownership); and cf *R v Rice* [1963] 1 QB 857, below.

ing a false declaration in an import entry form concerning certain bags of seed. Evidence was admitted that the bags of seed bore the words ‘Produce of Morocco’. The Privy Council held that the evidence was inadmissible hearsay and advised that the conviction be quashed. The decision may be usefully compared with that in *R v Lydon*.²² The appellant, Sean Lydon, was convicted of robbery. His defence was one of alibi. About one mile from the scene of the robbery, on the verge of the road which the getaway car had followed, were found a gun and, nearby, two pieces of rolled paper on which someone had written ‘Sean rules’ and ‘Sean rules 85’. Ink of similar appearance and composition to that on the paper was found on the gun barrel. The Court of Appeal held that evidence relating to the pieces of paper had been properly admitted as circumstantial evidence: if the jury were satisfied that the gun was used in the robbery and that the pieces of paper were linked to the gun, the references to Sean could be a fact which would fit in with the appellant having committed the offence. The references were not hearsay because they involved no assertion as to the truth of the contents of the pieces of paper: they were not tendered to show that Sean ruled anything.²³

In Steven Uglow’s book [102], we find his account of the same case.

“*written statements*: the classic case here is *Myers v DPP* ([1964] 2 All E.R. 877) where the defendant bought wrecked cars for their registration certificates. He would then steal a similar car and alter it to fit the details in the document. He would sell the disguised stolen car along with the genuine log book of the wrecked car. The prosecution sought to show that the cars and registration documents did not match up by reference to the engine block numbers and introduced microfilm evidence kept by the manufacturer, showing that this block number did not belong in a car of this registration date. The microfilm was prepared from cards which

²²[1987] Crim LR 407, CA.

²³See also *R v McIntosh* [1992] Crim LR 651, CA (calculations as to the purchase and sale prices of 12 oz of an unnamed commodity, not in M’s handwriting but found concealed in the chimney of a house where he had been living, admissible as circumstantial evidence tending to connect him with drug-related offences); and cf *R v Horne* [1992] Crim LR 304, CA (documents of unknown authorship, referring to H, containing calculations possibly relating to the cost of importing drugs, and found in the flat of a co-accused to which H was supposed to deliver the drugs, inadmissible against H). *R v McIntosh* was applied in *Roberts v DP* [1994] Crim LR 926, DC: documents found at R’s offices and home, including repair and gas bills and other accounts relating to certain premises, were admissible as circumstantial evidence linking R with those premises, on charges of assisting in the management of a brothel and running a massage parlour without a licence.

were themselves prepared by workers on the assembly line. Lord Reid in the House of Lords held that the microfilm was inadmissible since it contained the out-of-court assertions by unidentified workers.”

The labelled structure of the above is as follows.

Let

- $t : C$ The numbers assigned to the cars by the manufacturers are x_1, x_2, \dots
- $t' : C'$ The numbers in the cars' logbook are y_1, y_2, \dots

If $x_i \neq y_i$, then we get:

- $t + t' : C'' =$ the numbers on the cars and numbers on the registration documents do not match

where

- $t =$ description of how the microfilm supporting C was obtained and compiled.
- $t' =$ the cars' logbooks.

The candidate item of data for admissibility is

- $t : C$.

The following passage is Lord Reid's argument that $t : C$ should be inadmissible, i.e., Lord Reid wants to argue that t should also contain the phrase “do not use me”.

This is done in the logic of the labels. In other words, Lord Reid's argument has to do with the data inside t .

Here is Lord Reid's argument (technically it is part of t). It also quotes the arguments given in favour of admitting $t : C$.

***Myers v DPP* [1964] 2 All E.R. 877 at 886b–887h, per Lord Reid**

It is not disputed before your Lordships that to admit these records is to admit hearsay. They only tend to prove that a particular car bore a particular number when it was assembled if the jury were entitled to infer that the entries were accurate, at least in the main; and the entries on the cards were assertions by the unidentifiable men who made them that they had entered numbers which they had seen on the cars.

Counsel for the respondents were unable to adduce any reported case or any textbook as direct authority for their submission. Only four reasons for their submission were put forward. It was said that evidence of this kind is in practice admitted at least at the Central Criminal Court. Then it was argued that a judge has a discretion to admit such evidence. Then the reasons given in the Court of Criminal Appeal were relied on. And lastly it was said with truth that common sense rebels against the rejection of this evidence.

At the trial counsel for the prosecution sought to support the existing practice of admitting such records, if produced by the persons in charge of them, by arguing that they were not adduced to prove the truth of the recorded particulars but only to prove that they were records kept in the normal course of business. Counsel for the accused then asked the very pertinent question — if they were not intended to prove the truth of the entries, what were they intended to prove? I ask what the jury would infer from them: obviously that they were probably true records. If they were not capable of supporting an inference that they were probably true records, then I do not see what probative value they could have, and their admission was bound to mislead the jury.

The first reason given by the Court of Criminal Appeal for sustaining the admission of the records was that, although the records might not be evidence standing by themselves, they could be used to corroborate the evidence of other witnesses.²⁴ I regret to say that I have great difficulty in understanding that . . . Unless the jury were entitled to regard them, I can see no reason why they should only become admissible evidence after some witnesses have identified the cars for different reasons . . .²⁵

At the end of their judgement, the Court of Criminal Appeal gave a different reason. ‘In our view the admission of such evidence does not infringe the hearsay rule because its probative value does not depend upon the credit of an unidentified person but rather on the circumstances in which the record is maintained and the inherent probability that it will be correct rather than incorrect.’ That, if I may say so, is undeniable as a matter of common sense. But can it be reconciled with the existing law?

²⁴This is our footnote. “corroborate evidence of other witnesses” means in our LDS language “help with the flattening process”.

²⁵Our footnote: i.e., $u_1 : X$ is admissible only if some other $u_2 : X$ is already admissible. See objection $s_{3,2}$ below. LDS allows formally for putting item $u_1 : X$ in the database in such a way that it can be used only in the flattening process to support other items but not in deduction.

I need not discuss the question on general lines because I think that this ground is quite inconsistent with the established rule regarding public records. Public records are *prima facie* evidence of the fact which they contain but it is quite clear that a record is not a public record within the scope of that rule unless it is open to inspection by at least a section of the public. Unless we are to alter that rule how can we possibly say that a private record not open to public inspection can be *prima facie* evidence of the truth of its contents? I would agree that it is quite unreasonable to refuse to accept as *prima facie* evidence a record obviously well kept by public officers and proved never to have been discovered to contain a wrong entry though frequently consulted by officials, merely because it is not open to inspection. But that is settled law. This seems to me to be a good example of the wide repercussions which would follow if we accepted the judgement of the Court of Criminal Appeal. I must therefore regretfully decline to accept this reason as correct in law.

In argument, the Solicitor-General maintained that, although the general rule may be against the admission of private records to prove the truth of entries in them, the trial judge has a discretion to admit a record in a particular case if satisfied that it is trustworthy and that justice requires its admission. That appears to me to be contrary to the whole framework of the existing law. It is true that a judge has a discretion to exclude legally admissible evidence if justice so requires, but it is a very different thing to say that he has a discretion to admit legally inadmissible evidence. The whole development of the exceptions to the hearsay rule is based on the determination of certain classes of evidence as admissible or inadmissible and not on the apparent credibility of particular evidence tendered. No matter how cogent particular evidence may seem to be, unless it comes within a class which is admissible, it is excluded. Half a dozen witnesses may offer to prove that they heard two men of high character who cannot now be found discuss in detail the fact now in issue and agree on a credible account of it, but that evidence would not be admitted although it might be by far the best evidence available.

It was admitted in argument before your Lordships that not every private record would be admissible. If challenged it would be necessary to prove in some way that it had proved to be reliable, before the judge would allow it to be put before the jury. And I think that some such limitation must be implicit in the last reason given by the Court of Criminal Appeal. I see no objection to a judge having a discretion of this kind though it

might be awkward in a civil case; but it appears to me to be an innovation on the existing law which decides inadmissibility by categories and not by apparent trustworthiness . . .

Structure of Lord Reid's argument

$\Delta_1 : N =$ number on car A is a , (when assembled), and Δ_1 is the support of this claim.

$\Delta_1 =$ description of procedures of entering numbers during assembly.

We also have a common sense metalevel persistence principle: numbers on cars persist (don't fade away or change).

$$N \rightarrow \mathbf{Always} N.$$

Thus, according to Lord Reid, t is equal to:

$$t = \{\Delta_1 : N, N \rightarrow \mathbf{Always} N\}.$$

He wants to block the use of t by attacking the admissibility of Δ_1 .

Four reasons were quoted for the admissibility of Δ_1 and three reasons for non-admissibility:

r_1 : Evidence of this kind is admitted in Central Criminal Court.

r_2 : Judge has discretion to admit such evidence.

r_3 : This is a list of reasons given in Court of Criminal Appeal, namely:

$r_{3,1}$: The records were produced to show that the records were kept in the normal course of business (but not to prove the truth of the recorded particulars).

$r_{3,2}$: Although the record may not be evidence by themselves, they may be used to corroborate other evidence.

$r_{3,3}$: We do not have dependency on the credit of an unidentified person but rather on a probably reliable process of record maintenance, and can therefore admit them.

r_4 : Common sense rebels against rejection of such evidence.

s_0 : No reported case or any textbook as direct authority for admission.

It seems at this point that r_1-r_4 are stronger than s_0 .²⁶ So Lord Reid is trying to weaken the force of r_3 and r_2 by attacking them logically with s_3 and s_2 :

- s_2 : Judges do not have the discretion to admit legally inadmissible evidence.
- s_3 : Counter argument to r_3 comprising of:
- $s_{3,1}$: If the records are not intended to prove the truth of their entries, what are they intended to prove? (I.e., they are irrelevant!)
 - $s_{3,2}$: Either the records are admissible or not. There is no sense in which they can become admissible only after some other evidence to the same conclusion becomes admissible (see Footnote 25).
 - $s_{3,3}$: Such records are not public records which are admissible for reasons that they are open to the public for inspection and correction. The current law therefore does not support their admissibility.

Figure 1 shows the form of t , where E = admit evidence or ‘use me’.

To strengthen his case (i.e., strengthen the overall labels for $\neg E$, Lord Reid is attacking the label r_3 by putting forward $s_{3,1}$, $s_{3,2}$ and $s_{3,3}$. Note that the reasoning in the different boxes can be of different kinds!

Note that one of the points Lord Reid is making is s_2 , namely that trial judges do not have discretion to ‘admit legally inadmissible evidence’.

Compare this with the Barclays Bank example. So the force of the argument is to influence the flattening process: we have $r_1-r_4 : E$ and $s_0, s_2, s_3 : \neg E$, which one wins?

In this case the evidence was not admitted.²⁷

Uglove continues:

The House of Lords recognized the absurdity of their position but felt strongly that it was for the legislature to reform the law and create new exceptions. Parliament dealt with the problem of documentary hearsay with the Criminal Evidence Act 1965 which created an exception for trade and business records This was later extended by section 68 of the

²⁶In other words, it seems that a reasonable flattening process, weighing $\{r_1, r_2, r_3, r_4\}$ against $\{s_0\}$ will decide in favour of the former and thus admit the records. Note that no rules are given at this stage of how the decision is made. In some logics, where labels are confidence numbers, we can give a rule; e.g. admit iff $r_1 + r_2 + r_3 + r_4 > s_0$, but not here.

²⁷This decision was made by vote as described in the quote from [72] on our page 17.

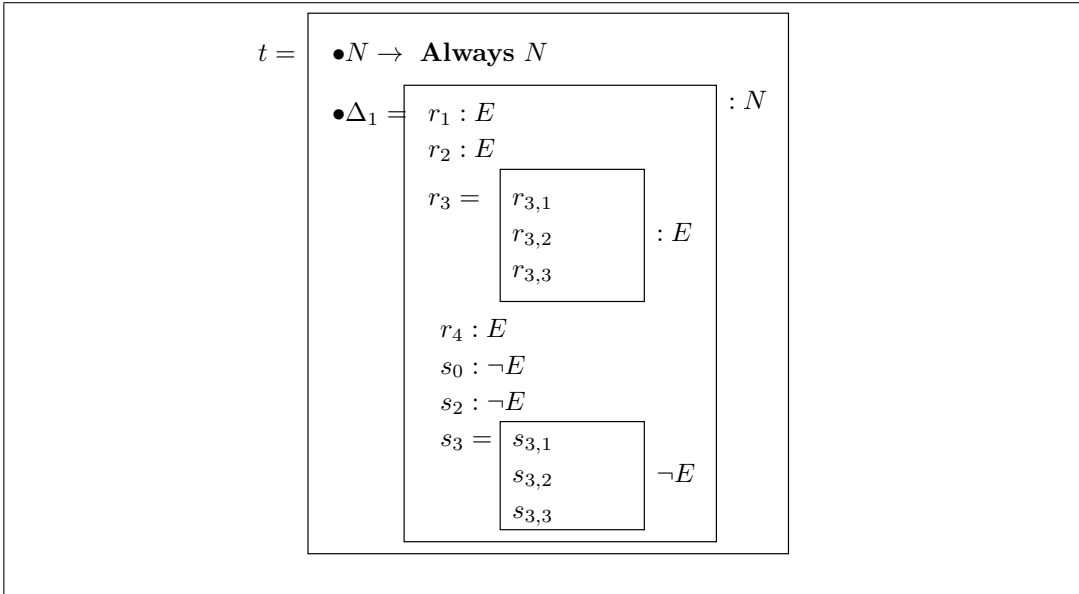


Figure 1

Police and Criminal Evidence Act 1984 and now by sections 23 and 24 of the Criminal Justice Act 1988. Such records have all been admissible in civil proceedings since the Civil Evidence Act 1968.

Myers has been regularly followed in such cases as *Patel v Comptroller of Customs* ([1965] 3 All E.R. 593) where the appellant was convicted of making a false declaration to customs, having stated that the bags of seed were originally from India. The prosecution sought to prove that the seed originated in Morocco and adduced evidence that the bags were stamped with ‘Produce of Morocco’. The Privy Council, following *Myers* held that these words were hearsay and inadmissible. Unlike *Myers*, there was no evidence that the writing was at all reliable, there being no testimony as to how or by whom the bags were marked.”

The reader should note that the main thrust of the argument and logic of the Lord Reid example is in weakening and strengthening labels. Put schematically we have a master argument, say E which can prove a conclusion on D . E is a labelled argument containing various labels within labels. Among this maze of labels there is a label t containing another argument, say Δ . To attack E we can attack Δ . Our argument attacking Δ can itself be attacked by attacking some label s in it and so on. This is reminiscent of systems of abstract argumentation theory. Bench-Capon

[9] has a paper on graphs of arguments and counterarguments, but his model is schematic. We can give actual proof rules and labelling disciplines so that questions like export from one label to another can also be considered. For example:

“If you weaken t then D will not follow from E , and that would be a bad precedent.”

One cannot argue in this way unless a specific labelled model is available. We shall examine the Bench-Capon paper in the next section. For the time being, we think that we have seen enough to be convinced that labelling logics can play a central role here, though we would understand if the cautious reader would prefer to reserve judgement until more case studies are presented.

4 Case Study. Sex Offender Case, Risk Assessment

Consider the argument structure of item Δ_1 of Figure 1. This can also be represented as a tree, as in

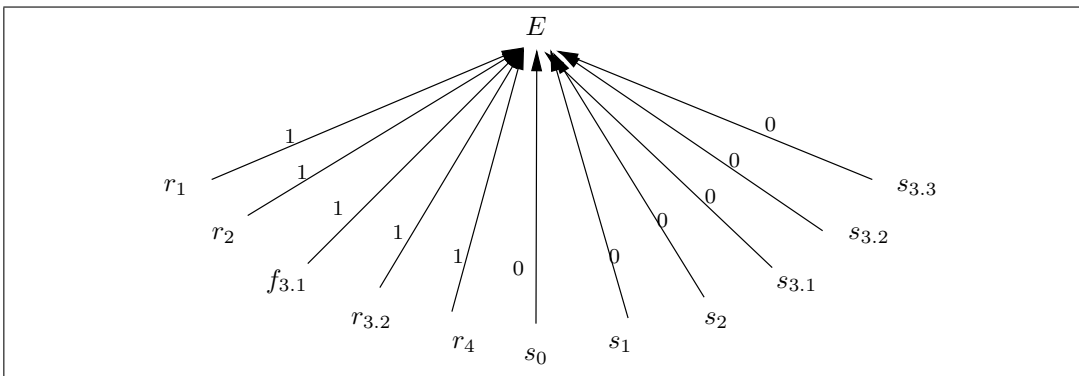


Figure 2

where $x \xrightarrow{1} E$ means x supports E and $y \xrightarrow{0} E$ means y supports $\neg E$ (i.e., $y \xrightarrow{1} \neg E$).

Figure 2 is the same representation as Figure 2 presented as a different geometrical form.

Now imagine a legal reporter interviewing Lord Reid and asking him a question about each of his arguments x and Lord Reid gives an answer. We can add to the tree the respective question and the answer. We can write $a(x)$ for the question about x and write $b(x)$ for the answer to $a(x)$. The reporter can also ask Lord Reid to provide a strength number $m(x)$ for each argument x . If this is done we get a graph like Figure 3.

In fact the legal reporter (who is most likely a legal man himself) might prepare

his interview and ask Lord Reid why he did not use certain other arguments which can maybe support E or support $\neg E$.

So we can assume, if we want, an additional set of arguments, say t_1, \dots, t_n , which could be relevant and Lord Reid has not mentioned.

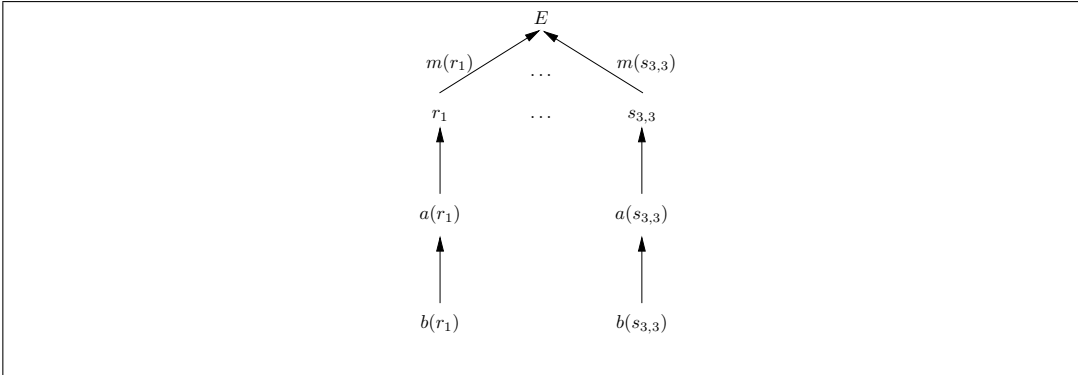


Figure 3

The above discussion connects the structure of Δ_1 of Figure 3 with the formal structure of our present case study about sex offenders risk assessment.

When a sex offender is convicted he/she can join a therapy group in prison (a group of 12–15 other offenders and benefit from therapy for several years, see [48]). He can also apply on the basis of his performance for a “good progress/behaviour” early release from prison. The decision is taken by a Judge, in consultation with a recognised expert in sex-offences, and it hinges on whether and how much risk there is to others in releasing the offender. The court deliberations are stylised and are structured as in Figure 3.

First the expert presents his qualifications and his report to the court and the sex offender’s lawyer can ask questions and he expert answers. Then the expert’s report is discussed. There are a fixed number of factors for and against:

E = this sex offender presents a risk to society.

The expert has to address each of them and give a risk number. The expert may have mentioned only some of the factors and then he may be asked why he did not mention other factors. This would be similar to the legal reporter asking Lord Reid why he did not mention some other relevant argument. The difference is in that in the sex offenders case there is an internationally recognized list of factors, while in the reporter case , he compiled a list himself. The lawyer can ask a question $a(x)$ about each factor x (i.e., what the expert assessed of x) and the expert answers $b(x)$.

The factors are fixed and there are international statistical packages giving a risk number $m(x)$ for each factor. These are updated all the time following international

case studies data.

In this case study we give a sample of x , $m(x)$, $a(x)$ and $b(x)$ written by Dr G. Rozenberg and based on 15 years' court experience of his court cases (see [48]). For the purpose of this article the data can be considered almost as transcripts of a realistic court case.

Let us begin:

The following is almost an actual court case created by Dr G. Rozenberg based on the many court cases in which he participated. It lists the questions he was asked and his answers. The text presented is a translation of transcriptions taken by the official court clerk of the Military Court in Jaffa, Israel, in a court cases of a certain sex offenders. The expert witness is Dr Rozenberg. The text is the exact detail of the cross examination questions presented to Dr Rozenberg and his answers. The original is in Hebrew and it was faithfully and accurately translated by the authors of [48]. Although their translation is not a legally accepted notarised "authorised translation", it is sufficient for the purpose of this article. Some slight modifications were made to avoid the possibility of identifying any offender.

It is a case study of an attack on the qualifications of the expert. It also has a stylised structure. It is comprised from questions about how the expert acquired/studied for his qualifications, as well as how he went about (methodology used in) composing his report. There are also questions about specific items in his report, not with a view of attacking the item but with a view to see if the expert understands their significance. Other questions relate to whether the expert understands the limitation and margins of error of his report.

There are also traditional Fallacies, trick questions, tempting the expert to answer in such a way that he appears to be racist, over-confident and full of his own importance, unable to take criticism or timid, hesitant and unsure of himself. We must remember that these questions are asked on the witness stand in front of a Judge and are intended to discredit the expert.²⁸

We shall not present the discussion of the qualification of the expert (see [48], but reproduce here the discussion of his expert testimony.

5 The Attack on the Expert Testimony

5.1 Background

There is consensus in the international community (ATSA — Association for the Treatment of Sexual Abusers) on the factors which contribute to the assessment of

²⁸Doug Walton and John Woods jointly and separately have written many wonderful books on the Fallacies. The reader should consult the internet.

risk of sexual offenders. There are also several actuarial tools to help the expert in assessing the risk of a given patient. The tool asks the expert to evaluate/answer questions about the individual patient and then gives a risk assessment a final grade which is a number x , $k < x < m$, where x, m, n are integers (depending on the tool). The expert can use several tools, as well as some additional clinical factors (determined by the experience of the individual expert) and the expert integrates all these results (in his own mind, as there is no Super Integrating Tool) into a final determination.

There is no super-tool which can integrate/reconcile the results of several existing tools. The expert has to decide which tools to use and how to integrate them. The ATSA list of factors are recognised by the Israeli courts, and the expert witness is expected in court to address these factors and be challenged by the defence attorney of the sex offender. The main tools are listed in the Appendix and the typical questions and answers in court are presented in this section. The list in this section does not represent any particular court case but is based on 11 years practice and thousands of expert opinions put forward by the second author. The courts follow Israeli law. The defence lawyer may invite his own expert to present a possibly different report and different conclusion. In this case the second expert will also appear in court and be subjected to the same procedures as the first expert with the prosecutor performing the attacks on the second expert factors.

This section deals with the attacks on the expert testimony. The structure of the testimony is as outlined in the previous section. Each numbered item below represents an attack sequence on a factor s . Each item comprises of three sub-items;

- what the expert says concerning factor i , denoted by s_i . (The expert can either introduce the factor in his considerations or not mention it at all. The factor may support the increasing of risk assessment of the offender or support the decreasing of the risk assess of the offender. There are international packages which assess the contribution of such factors s_i .)
- the attack on what the expert says, denoted by a_i . (This attack is mounted by the defence. So if the expert does not mention a factor which decreases the risk assessment the defence can ask why? If the factor increases the assessment of risk the defence might add information which makes the increase smaller. If the expert gets his facts wrong, then his entire testimony is at risk and the expert loses credibility. So this does not happen in practice. Note further that the node a_i denotes all of what the defence says which can be comprised of several attacks in the formal sense, or a joint attack or a higher level attack, etc.)

- the experts answer to the attack denoted by b_i . (Many of the answers of the expert are explanations or more information, see [44].)

The factors come with labelling of strengths: low, moderate and strong. We shall see in the Appendix, which surveys Tools which assess the strength of these factors, that numerical strength are assigned to them both positive and negative numbers, the qualitative strengths can be derived from these numbers. Note that the factors s_i can be factors which increase risk or sometimes factors which decrease risk (such as participation in therapy). We still view them as “supports” with negative input, which turns them as “attacks”. The examples below show that the “counter attacks” a_i on s_i can either question the strength and the significance suggested by s_i or they can question the validity of s_i in applying or not applying to the sex offender in question or can be factual attacks on the factual part of the factor s . Some attacks a_i are logical fallacies. The replies b_i to the items a_i are more in the nature of explanations, rather than “counter-counter-attacks” on a_i .

We need to be more explicit here. Let us assume that the expert puts forward factor s_i . Factor s_i has two parts, the factual part and the assessment part arguing its contribution to how dangerous the offender is. For example The lawyer of the defence attacks s_i with counter argument a_i . If the counter argument is successful against the factual part, then the credibility of the expert is shattered, and all his support arguments s_j for all j are destroyed. Take s_{15} for example. The factual part is that the offence was in a public place. The attack $a_{15}(b)$ simply says that the attack was at night at an isolated part of the public place and so the factor should not be used. This is not a factual attack. But if the defence proves $a_{15}(a)$, that the attack was at home, then this is a factual attack and all the support arguments s_j for all j are destroyed. On the other hand if the lawyer’s attack $a_{15}(a)$ is factually destroyed by b_{15} , then his other arguments can still be used. The lawyer is not an expert, he is not committed to the same credibility criteria, and he is expected to try all kinds of arguments. a_{15} may be destroyed but his other arguments may survive. The defence lawyer may invite his own expert to present a possibly different report and different conclusion. In this case the second expert will also appear in court and be subjected to the same procedures as the first expert with the prosecutor performing the attacks on the second expert factors.

We finally would like to put the contents of this section (namely the attack on the expert testimony) into a general perspective from the point of view of argumentation: An influential classification of dialogue types is that of Walton and Krabbe [112]. We recall their distinction between persuasion and deliberation dialogue. The goal of a deliberation dialogue is to solve a problem while the goal of a persuasion dialogue is to test whether a claim is acceptable The material of this section falls under the

category of persuasion dialogues. In such dialogues, two or more participants try to resolve a difference of opinion by arguing about the tenability of a claim, (in our case the degree of risk of a given sex offender), each trying to persuade the other participants (in our case mainly the Judge) to adopt their point of view. General dialogue systems regulate such things as the preconditions and effects of speech acts, including their effects on the commitments of the participants, as well as criteria for terminating the dialogue and determining its outcome. Good dialogue systems regulate all this in such a way that conflicting viewpoints can be resolved in a way that is both fair and effective [79]. In our case the procedure as we described is a highly stylised tree of depth 4, and the final arbiter is the Judge.

Furthermore the particular arguments used are informational and numerical, as we shall see in later sections.

The reader would also benefit greatly from looking at the important paper of Gordon, Prakken and Walton, [64] and the survey [19].

Let us begin.

Full Matrix/List of Relevant parameters/ factors to assess sexual risk

Note that the attacks on these factors are taken from protocols of actual cases involving Dr Rozenberg and his actual replies. They are not from a single court case but a representative compilation. But each sequence was actually asked and answered in court. The wording describing the node s_i is the authors wording simply saying the factor was or was not introduced in the experts report. We could have written “+” and “-” . The entries for a_i and b_i are from transcripts of actual court cases.

5.2 This factor is the age

Sex offender’s age taken into account when making the risk assessment. Below is the official table of the age groups and the risk strength assigned to them

Risk factor	Age group
1	18–34.9
0	35–39.9
-1	40–59.9
-2	60 or older

A significant factor with at least moderate importance

- s_1 The expert gives a contribution due to this age factor

- a_1 The attack says that the offender is older so according to the table the risk factor strength should be less.
- b_1 The expert reply: Recent literature shows the relationship between the age of the offender to a level of sexual risk is not so dichotomous, for example, we learn that the dangerousness decline in child molesters is milder and occurs in older ages than among rapists. Also, the person who committed the offence in an advanced age, his age should not be taken that seriously as a risk reducing factor.

5.3 Division/ classification of sex offenders by the official definition of the nature of their offence

child molester- victim under age 13

rapist- victim above 13 years old. A significant factor with at moderate importance

s_2 — The expert gives a contribution of risk due to this factor.

a_2 — The attack says that the expert should have taken into account that risk of rapists against the passage of time declines at a faster rate than that of in child molester.

b_2 — Expert reply: Recent literature shows the relationship between the age of the offender to a level of sexual risk is not so dichotomous, for example, we learn that a dangerousness decline in child molesters is milder and occurs in older ages than among rapists. Also, the person who committed the offence in an advanced age, his age as a factor that reduces dangerousness should be taken with a grain of salt.

5.4 Family status

The official classification is as follows:

Bachelor — a person who has not lived with an Intimate Partner nor had a joint household with a partner for a period of at least 2 Years. If bachelor then this factor raises the dangerousness.

This factor is of Low importance.

s_3 — The expert put forward this factor

a_3 — The attack: You can see that the accused person is acquainted with a woman, maybe even married her, and managed a relationship for almost 2 years. Technically he is considered a bachelor but arguably it teaches us about his capabilities and reduces risk.

b_3 — Expert reply: The literature indicates that the fact a person contacted and possibly married is insufficient. Only if he would be able to manage relationships

with common household for two years it will show the ability to keep significant relationship.

5.5 Index Non-sexual Violence (NSV) - Any Convictions

If the offender's criminal record shows a separate conviction for a non-sexual violent offence at the same time they were convicted of their Index Offence, this factor raise the dangerousness.

A significant factor with at least moderate importance

s_4 — Expert mentions use of violence.

a_4 — The attack: If the offender's criminal record does not show a separate conviction for a non-sexual violent offence at the same time they were convicted of their Index Offence, this factor should be ignored.

b_4 — Expert Reply: Do not ignore the fact that almost all sex offences include aspects of coercion and violence and the choice to convict a person of a crime of violence is a legal issue rather than sex offence issue.

5.6 Prior Non-sexual Violence - Any Convictions

Having a history of violence is a predictive factor for future violence. A significant factor with moderate importance

s_5 — Expert did not address this factor, (meaning that in the court case this factor was not mentioned in the expert's report. Since this is a mitigating factor the defence asks why was it not mentioned).

a_5 — The attack: If not convicted, so arguably he usually keeps the law and it is one-time lapse and the current conviction probably discourages him.

b_5 — Expert reply: Sometimes the person tells us himself that once he used violence against family members or others and the absence of conviction of violence does not necessarily indicate that he never used violence.

5.7 Prior Sex Offences

The best predictor of future behaviour, is past behaviour. A meta-analytic review of the literature indicates that having prior sex offences is a predictive factor for sexual recidivism.

A significant factor with high importance

s_6 — expert mentioned that the person had previous offences which increase the risk.

a_6 — Attack : This was a long time ago. Since then for many years there were no conviction. So previous conviction probably discouraged him.

*b*₆ — Expert Reply. Criminal that have several conviction at any time in the past is still to be considered dangerous. The existence of a conviction for sex offence often indicates quality of functioning of law enforcement officials and victims readiness and motivation, (if such were indeed), to complain. Also, in law, sometimes for a similar offence the offender can be convicted on different offences, for example, reveals himself in public might be convicted of committing a public indecent assault, but charges may be ether wild behaviour in a public place.

5.8 Prior Sentencing Dates

This item relate to criminal history and the measurement of persistence of criminal activity. The Basic Rule: If the offender's criminal record indicates four or more separate sentencing dates prior to the Index Offence, the offender is more dangerous. Count the number of distinct occasions on which the offender was sentenced for criminal offences. The number of charges/convictions does not matter, only the number of sentencing dates.

A significant factor is law importance

*s*₇ — Expert used this factor, even though the past convictions were not sex related.

*a*₇— Attack: If not convicted before, so arguably he usually keeps the law and it is one-time lapse and the current conviction probably discourages him. We can claim that if the subject made prior offences that teach about his criminal lifestyle, the risk sex assessment should evaluate only sexually dangerous and nothing else and the index offence is one-time lapse. People with criminal life style mostly feel disgusted by sex offences and shy away of it and their self-esteem injured therefore current conviction probably discourages him

*b*₇ — Expert Reply: A person who has a background of criminal offences shows difficulty to maintain limits and respect the boundaries of correct behaviour and one of the main concerns is that reluctance not to respect the laws and other limits may result in repeated sex offences, too.

5.9 Any Convictions for Non-contact Sex Offences

Offenders with paraphilic interests are at increased risk for sexual recidivism. Offenders who engage in these types of behaviours are more likely to have problems conforming their sexual behaviour to conventional standards than offenders who have no interest in paraphilic activities. If the offender's criminal record indicates a separate conviction for a non-contact sexual offence, the offender is more dangerous.

A significant factor with high or very high importance

s_8 — Expert did use this factor

a_8 — Attack: You can argue that sex is contactless low threshold of severity of injury and despite the offence with high recidivism, even if a person carries the offence again, the damage it can cause to the potential victim not so strong a man performing very offensive offence with contact and entering offences. Typically, offenders who committed Non-contact Sex Offences contact offences are less likely to make contact sex offences.

b_8 — Reply: The person that makes risk sex assessment is not a judge, and it is not his job to determine severity of harm, but to indicate to which group the subject belongs and what are the chances that he will make again sex offences, regardless of the severity of the offence.

5.10 Unrelated Victims (victim known to the offender, but not family)

The items concerning victim characteristics. Sex offence on Unrelated Victims related to higher risk assessment. Research indicates that offenders who offend only against family members recidivate at a lower rate compared to those who have victims outside of their immediate family.

A significant factor with high importance

s_9 — Expert used this factor

a_9 — Attack: Offender who harm the victims in his family is less dangerous because he is often perceived as a "lazy" who probably will not look for victims outside the family.

b_9 — Reply: Despite the fact that the person who harm victims within the family hurts somebody outside the family is relatively low, but it still exists. In addition, the offence to be possible because of problematic family climate expressed within weak limits and if the family circumstances do not change, significant treatment, then the individual may return to the same environment that allowed the violation in the past and may again exploit his authority and hurt.

5.11 Any Stranger Victims?

The Basic Principle: Research shows that having a stranger victim is related to sexual recidivism. If the offender has victims of sexual offences who were strangers at the time of the offence (stranger is defined as a person known to offender for less than 24 hours prior to the offence), is related to higher sexual recidivism.

A significant factor with high importance

s_{10} — Expert says the victim was a stranger.

a_{10} — Attack: A strong connection formed between the offender and the victim, even though they met less than 24 hours (they had intimate conversation before the offence).

b_{10} — Reply: But he hurt the victim, who is not a relative and possibly in future is pushing a minimal introduction to compromise.

5.12 Any Male Victims?

The Basic Principle: Research shows that offenders who have offended against male children or male adult recidivate at a higher rate compared to those who do not have male victims.

A significant factor with high importance

s_{11} — The expert used this factor

a_{11} — attack- you say that a sex offender attacking male victims is more dangerous than offender who attacks female victims. This is clearly a prejudiced judgement between males and females. You see a man attacking another man as sick and therefore you make him more dangerous.

b_{11} — Reply There is no prejudice here, the observation is based on statistical data.

5.13 Alcohol consumption is clearly associated with violence

This is a strong factor in assessing risk. s_{12} — The expert increased the risk owing to the offender's high alcohol consumption

a_{12} — Attack 1: the offender has rehabilitated, he is no longer drinking.

$a_{12}(b)$ — Attack 2: the man has been alcoholic for a long time without offending, so there is no real connection.

b_{12} — Reply: The expert assertion about use of alcohol is based on the offender report of his use of alcohol, and it is well known that such reports can be unreliable. The offender report of alcoholism could be a cover for some more serious pathological causes.

$b_{12}(b)$ — Furthermore the use of alcohol can cause offence while drunk. This is a worrying factor because he might drink and be inhibited in the future and offend again.

5.14 The use of hard drugs

The connection between being a drug addict and sexual offence is not strong enough. Research identifies two types of drugs (excluding alcohol) contribute to hyper sexuality, namely Cocaine and Meta-amphetamines. To the extent that we get confirming

scientific reports about the connection, we will consider drug abuse as a risk factor. At any rate this is a weak factor

s_{13} — The expert mentions this as a factor.

a_{13} - Attack 1 —The offender has rehabilitated, he is no longer drug addict.

$a_{13}(b)$ — Attack 2 The man has been addict for a long time without offending so there is no real connection.

b_{13} — Reply. The expert assertion about use of drugs is based on the offenders report of his use of drugs, and it is well known that such reports can be unreliable. The offender report of drug addiction could be a cover for some more serious pathological causes. Furthermore the use of drugs can cause inhibited behaviour and to lead to offence while under the influence. This is a worrying factor because he might use drugs in the future and offend again. Note that meta-amphetamines do increase /flood the sex drives and therefore might push the man to further offence.

5.15 Sexual offence while the offender was under court order

This could be, for example, a legal trial, conditional sentence, legal restrictions, etc. This is a strong factor

s_{14} — Expert used this factor.

a_{14} — Attack - the offender has been punished and will behave. Furthermore he did not understand at the time the full meaning of legal restrictions but now he does understand.

b_{14} — Reply: Maybe the offender just says he will now behave but this does not ensure that he will not offend again.

Furthermore the effects of the present trial and punishment will wear off as time goes by.

5.16 Sexual offence in a public place

This is a medium strength factor

s_{15} — The expert used this factor.

$a_{15}(b)$ — Attack- The offender made his offence at night at insulated place and the chance that somebody would see him is low.

b_{15} — It is still a public place and even at insulated places people can pass. It is known that offending in a public place indicates a deep difficulty to restrain oneself and control one's drives.²⁹

²⁹One of the referees made the following comment about this case (factor s_{15}), I quote:

“Why might someone not attack on the basis that it was raining, so there was a lower chance of being interrupted? Or in a place that was not visible to passers-by? Why

5.17 The use of force while offending

This includes using firearms or the threat of using firearms, or use of physical force, or threat of physical damage or kidnapping.

This is a medium strength factor.

s_{16} — Expert uses this factor.

a_{16} — Threat is not really use of force.

b_{16} — professional literature shows it is it. Threat is definitely count as a use of force. Many times it is enough to compel person to make things that he didn't. Conviction of violence in addition to conviction of sexual offence indicates the offender not only cannot control his sexual drives but also cannot control his aggression.

5.18 The offender subjected the victim to a variety of sexual violations

These include: Penis penetration to vagina, finger into vagina, foreign object into vagina, groping the victim, masturbating over the victim, forcing the victim to grope the offender, forcing victim to masturbate, Forcing victim to give offender oral sex, offender giving victim oral sex, offender exposes himself (excluding exposing for the purpose of executing the offence), forcing victim to make sex with a third party/object, penetration of penis to anus, penetration of finger to anus, penetration of object to anus, kiss, forcing the victim to masturbate the offender.

This is a weak factor

is the attack a conjunction of night time and isolation — surely isolation could be enough to form an attack? What I would expect is that the typical attacks would be evidenced through reference to the court record — and then I would expect to see many different attacks that might be levelled arranged into groups, classes, or hierarchies perhaps. Similarly with defences against those attacks."

We note that s_{15} is a transcript of a case in court. The reader might ask whether we have collected an exhaustive list of transcripts and analysed them and examined them? Maybe the above suggested referee questions were asked in other cases? The answer is we did not assemble a larger set of transcript but a representative one. There is sufficient data and we learnt a lot from these examples already, namely the idea of the attack as information input, see [44].

Let us examine the transcript of s_{15} itself, to show the reader what we mean by representative. The attack a_{15} adds factual information, and tries to say, given this information, then the place was not really public. The response b_{15} is actually saying that the factor's contribution to the risk assessment of the sex offender was determined statistically based on the formal definition of public place (as opposed to the concept of not containing people) and the extra information is not relevant to the statistics. Again b_{15} is an attack by adding information.

In fact b_{15} is also a valid counter-attack to the referees suggestions above ("it was raining", or "it was in a place that was not visible to passers-by", etc...), again because such cases did not go into the statistics!. Compare with b_{20} .

s_{17} — Expert lists the offences done by the offender

a_{17} — attack. These should be considered a single offence and not a list of multiple offences. Moreover, almost any rape or other sex offence including a variety of sexual violations. For example, it is almost impossible to rape without groping the victim.

b_{17} — Reply: yes legally it is a single offence, but statistics shows that multiple components increase risk of re-offending in the future. The offender needs multiple stimulations to satisfy his drive. The offender might even commit some unusual acts in the future, and if the indictment detail the violation, than probably is was a different offence and not a basis to perform another offence.

5.19 Sex offender with victims from different age groups

In such a case the offender is considered more dangerous because the offender has a larger group of potential victims.

The age groups are:

0–6.99; 7–12.99; 13–15.99; 16 and above

s_{18} — Expert mentions this factor

a_{18} — Victims may not look their ages so it only an illusion that the offender is not focused on a single age group.

b_{18} — Reply. As an expert I have a choice and judgement on whether I work like a simple mathematical machine or try to decide on the correct evaluation and scenario. I try to understand the triggers motivating the offence and using that evaluate how dangerous the offender is and to what age groups. I especially examine the significance of cases where the victim's age is near the boundaries.

5.20 Age of victim is 13–15 years

An offender attacking this age group is more dangerous if the offender is 5 years older or more than the victim.

This is a medium factor

s_{19} — Expert mentions this factor

a_{19} — Attack. The age division into group is arbitrary and further teenagers.

Vary in how old they look, and many times 13–15 years old looks like elder.

b_{19} — Reply: the expert exercises judgement. The problem here is that the offender seeks an intermediate age group between children and grownups. There is the danger of a shift into the neighboring age groups. It is offenders responsibility to know the exact age of teenager. And mostly the confusion is a result of cognitive distortion of the offender.

5.21 Offender has not been able to maintain continuous employment up to the offence

This is a medium factor

s_{20} — Expert quotes this factor

a_{20} — There is an objective market difficulty in maintaining continuous employment. Many employers sack people in order not to give them tenure.

b_{20} — This is a statistical observation. The statistics show increase in risk. The statistics does not consider the reasons behind the lack of past continuous employment.

5.22 Offender violated some restrictions imposed by court orders, not necessarily sexually connected

This is a medium factor.

s_{21} — Expert mentions this factor

a_{21} — The past offences are not sexual, why are you mentioning them?

b_{21} — The offender cannot keep to proper boundaries, and his “internal policeman” is weak. If within the boundaries of court orders the offender could not police himself, he might reoffend if we release him now.

5.23 Empathy towards the victim

Weak factor

s_{22} — Expert mentions this factor

a_{22} — The literature shows there is no significant connection of this factor to risk.

b_{22} — If there is no empathy to the victim the offender will not appreciate the damage he is doing, and will not be interested or respond well to remedial treatment.

5.24 Disrespect to authority and institutions

s_{22} — expert mentions this aspect

a_{22} — The literature shows there is no significant connection of this factor to risk

b_{22} — If offender does not respect authority, then the offender if released with disrespect the officer supervising him/her and will try to out-manoeuvre the officer and offend again

5.25 Medical treatment to lower the sexual drive

This is an important factor, medium strength, as long as the patient participates

*s*₂₄ — expert mentions this.

*a*₂₄ — The Offender agrees to a chemical castration without being forced to do it. He is risking his body and might have to face side effects. This is a proof of how much he appreciates his wrong doing in the past and shows commitment to be risk free in the future. This must be considered a significant factor.

*b*₂₄ — This treatment affects the offender capabilities, not his personality and tendencies. Therefore without a genuine internal change there is still the risk of further offence, especially if the treatment is discontinued.

Furthermore the offender agreeing to the treatment may be just manipulative and not genuine, and we can be sure only if he continues with it for a considerable period of time. This is why this factor doesn't change the risk assessment in the long term..

5.26 No community or family support for the offender

Low factor.

*s*₂₅ — Expert mentions this factor

*a*₂₅ — Offender can take care of himself

*a*₂₅(b) — It is bad enough that everyone abandoned the offender, you have also to punish him for it?!

*b*₂₅ — This is not a punishment but the unfortunate fact that the offender will have no support to help him not offend again.

5.27 Offender is mentally retarded

This increases risk, medium factor.

*s*₂₆ — Expert mentioned this factor

*a*₂₆ — This is God's doing, what can the offender do?

*b*₂₆ — Mental retardation leads to dis-inhibition. The offender cannot learn from experience or appreciate vague situations with unclear boundaries.

5.28 Mental illness

Medium factor for increase in risk

*s*₂₇ — Expert mentions this factor

*a*₂₇ — What can he do, it is not his fault.

b_{27} — Mental illness leads to dis-inhibition. The offender has difficulties to learn from experience or appreciate vague situations with unclear boundaries.

We are not supposed to be politically correct but we deal in science and it is proven that mental illness increases risk of re-offending.

5.29 Offender does not accept responsibility for his actions nor expresses regret

Factor of low importance

s_{28} — Expert mentions this factor.

a_{28} — The literature does not consider this significant

b_{28} — If the offender does not accept responsibility of regret he will not be interested in any change. Accordingly, his chance to integrate on treatment and to derive the usefulness from it is low.

5.30 Did the offender plead guilty?

This is low factor.

s_{29} — Expert mentioned this factor

a_{29} — A literature do not attach much importance to this factor with the possible exception of a small group of offenders.

b_{29} — For offences within the family unit this is an important factor.

Furthermore, it is less likely the offender will accept treatment nor benefit from it

5.31 The offender has a distorted way of thinking

Low importance.

s_{30} — Expert mentions this factor

a_{30} — This factor is not identified in the literature. Besides, everyone has distorted ways of thinking one way or another.

b_{30} — Sex offenders have their own characteristic distortions, that form the basis to rationalise and justify his offences. We know there is a connection between thinking positions and behaviour.

5.32 Offender has low opinion of himself

Medium importance for increasing risk.

s_{31} — Expert presents this factor.

a_{31} — Person with low opinion of self the offender will not dare offend.

b_{31} — On the contrary offender will not dare approach normal relationship and will find someone weak to offend and attack.

5.33 Offender is physically or mentally impotent or is ashamed of his sexual organs

Factor of medium to high importance

s_{32} — Expert mentions this factor as increasing risk

a_{32} — On the contrary, there is no risk, he cannot do it he will not do it.

b_{32} — Not at all, we are dealing with frustration as a basis for action. To prove him-self the offender might prey on the weak such as children.

5.34 Impulsiveness, low tolerance to stimuli

Factor of medium importance.

s_{33} — Expert presents this factor

a_{33} — Usually his impulsiveness is not connected with sex

b_{33} — Impulsive people are unpredictable, you cannot be sure what the offender will do.

5.35 Strong sex drive

Factor of high importance for risk.

s_{34} - Expert presents this factor

a_{34} – So what, the offender will just be busy masturbate more often and is less likely to offend.

b_{34} – Research shows that on the contrary, increase masturbation enhances existing sex drives and not diminishes them. The offender is more likely to seek real contact.

5.36 Sexual deviation

Such as pedophilia, exhibitionism, proterism, etc.

Factor with high risk.

s_{35} — Expert uses this factor.

a_{35} — The man is sick, he needs hospital, not punishment.

b_{35} — I am not a Judge, the fact is that people with sexual deviation are high risk offenders.

5.37 Offender completed medical treatment

This is medium factor in reducing risk.

*s*₃₆ — Expert did not include this factor

*a*₃₆ — The offender did conclude a treatment why did you not include it as a high risk reducing factor?

*b*₃₆ — The treatment is not effective on some people. They emerge from it with some success but these fade in time. The real test is if the offender continues the program suggested by the treatment.

5.38 Sex offender treatment was interrupted and never completed

High risk factor.

*s*₃₇ — expert uses this factor.

*a*₃₇ — The interruption was due to objective factors such as the offender was sent to prison and was not allowed to complete the treatment.

*b*₃₇ — Even if it is not the offender's fault the fact is that half a treatment is risky and makes the situation worse in confusing the patient.

5.39 Does the offender understand/ know the risk/ trigger situations? Can the offender use adaptive preventive measures?

Medium factor

*s*₃₈ — The Expert said the offender did not know.

*a*₃₈ — The offender did know but when you talked to him he was under stress and could not list them. Anyway there is not enough research about this factor

*s*₃₈ — It is important to know the risk/ trigger situation for offence and learn to avoid them. It is important for the offender to know that even simple, seemingly unimportant decisions can put him at a risk of a trigger situation.

5.40 Personality disorder

Factor of low importance.

*s*₃₉ — Expert mentions this factor.

*a*₃₉ — There is not enough research on this factor.

*b*₃₉ — Sometimes this can be the reason for the offence. For example a narcissist might think the victim actually wants sex and the offender is actually being helpful.

Personality disorders are very difficult to treat.

5.41 The offender has had a long prison sentence

Factor with low strength.

s_{40} — Expert mentions this factor.

a_{40} — Offender did not offend in prison and suffered long enough. Why don't you let go instead of continuing to support punishing him?

b_{40} — I don't deal with punishment. I deal only with risk assessment. Today there is literature that indicates that having served a long prison sentence does not reduce risk but might even increase risk.

6 Value-based Argument Framework

The purpose of this section is to compare our approach with that outlined in Bench-Capon [9] and to show how labels can be used more effectively. We also give a Bayesian approach and a neural nets approach. In coming work we hope to offer an LDS mix of all approaches. We believe any realistic model needs to do that!

We can indicate at this stage how the abstract argumentation model can relate to LDS. Consider the Lord Reid argument as presented in Figure 1. It has arguments r_1, \dots, r_4 in favour of E and counter arguments s_0, \dots, s_3 in favour of $\neg E$, essentially attacking r_1, \dots, r_4 . LDS requires in this case a flattening function (or a process) to tell us which arguments win and at what strength we can use E or $\neg E$.

This flattening process can make use of abstract argumentation theory, either in its Bench-Capon form, or modified with probability or implemented in neural nets. A taste of these options is given in this section.

6.1 The Framework

We begin by discussing and highlighting our method of modelling. The first principle is to work bottom up from the application area into the formal model, trying to reflect in the formal model more and more key properties of the application area. In the case of evidence this means we need to see and study many examples/case studies/debates about evidence and then try to construct a suitable logic for it. Chances are that existing logics, constructed for some other purpose, may not be the most suitable. Our starting formal system for this purpose is LDS. The theory of LDS was developed from the bottom up point of view, especially to model aspects of human behaviour, reasoning and action, and is very comprehensive, adaptable and incremental. It contains a large variety of existing logical systems as special cases. What is more important is that LDS is not a single system but a methodology for

building *families* of systems, ready to be adapted to the needs of various application areas, in our case to the theory of evidence.

One very important side effect of this approach is that the logic can be worked up directly from the day-to-day activity of the practitioner of the laws of evidence, without necessarily forcing him to study logic. The ‘logic’ will be hidden in the stylised movements he will be asked to make, and the interplay between the labels and comments and arguments he will be using.³⁰

In contrast to our approach, in a good deal of applicational work in logic, a logic is applied to various areas and tend to force the application area into a form suitable for its existing formalism. This tends to produce results intelligible mainly to the logician, ignoring that the ordinary human/lawyer/judge already knows intuitively how to handle his daily life, and that all he needs is some bottom up additional organisation of his activities which will enable him to understand it better and possibly solve some of his outstanding puzzles.³¹

³⁰For recent work on the norms implicit and tacit in the cognitive behaviour of parties to a criminal trial, see Woods [118, chapter 20, ‘An epistemology for law’], and [120]. Consider the widespread use of anagrams. Take as an example the pair of words ‘read on’. We can rearrange the letters (including the space between the words) into ‘no dear’. Let us write this as

read on ⊢ no dear

We can also write equivalently

space, a, d, e, n, o, r ⊢ read on
 space, a, d, e, n, o, r ⊢ no dear

where on the left we just listed the basic blocks we can use, including the space.

Now suppose we allow you some ‘wildcard’ of the form

space ↦ any other already listed letter

Then we get

space, a, d, e, n, o, r, (space ↦ any other already listed letter) ⊢ adorned

We chose here space ↦ d.

What we have been doing here was linear logic!

So anagrams with wildcards is linear logic.

The idea that logic can be ‘translated’ into stylised proof movements was put forward in the Gabbay 1984 logic lectures at Imperial College, London. See the first chapter of [34] and see [36]. Peter Tillers says similar things in his paper in [80, pp. 2–11]. We assume the word ‘dynamics’ in the title of [80] is significant.

³¹The modelling practices of the social sciences generally are adaptations of the modelling paradigms of physics (rather than, say, biology), and are a reflection of the primacy of logical positivism as the social sciences were in process of articulating its philosophical presumptions. But

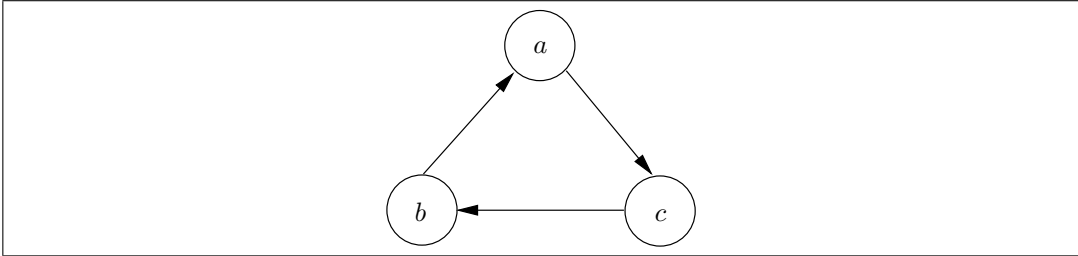


Figure 4

The difference in this point of view is apparent when we look again at Prakken's book. The book does realise the potential in the interaction of logic and law. It also recognises some of the kinds of logics needed to model some aspects of the law. But having made and argued all of these points, the main part of the book gives an exposition of the relevant parts of the logic in a way that only a logician can understand. This is also true at the moment of this version of our article, but we hope in the full version to be able to do logic directly in the legal evidence application area. See Footnote 30.

Having said all that, we can now look at some specific model, namely that of abstract argumentation systems. These were put forward as a response to the realisation that no argument or proof is conclusive in real life, and that arguments have counterarguments. The argument framework has the form $AF = (AR, Attacks)$ where AR is a set of objects called arguments and $Attacks$ is a binary relation (usually irreflexive), saying which arguments x attacks which argument y . The following Figure 4 is an example

a attacks c , c attacks b and b attacks a .

There are no winning arguments here. This framework is too abstract to be of specific use. It equally applies to circuits and impending circuits, credits and debits, neural nets and counterweights or any system involving x and anti- x , whatever x is.

To apply such a system successfully we need to go into the structure of the arguments and analyse the mechanics of one argument attacking another.

Bench-Capon tried to improve upon such systems by introducing a clever idea; the value-based argumentation framework. In this framework we are given a set of colours (values) and a colouring of the arguments. The values are partially ordered and an argument of strictly lesser value cannot now attack an argument of stronger value.

it is almost never satisfactory to abstract from the data of human interactions in the same way that one abstracts from the interactions of physical particles.

So following Bench-Capon in the previous figure, if we make b red and a and c blue then

1. If blue is stronger than red, then b cannot attack and defeat a , a can attack c and the winning arguments are $\{a, b\}$, because c is out.
2. If red is stronger than blue then the winning arguments are $\{b, c\}$.

Certainly this colouring with values is an intuitively welcome improvement. However, this model is still too abstract. Real life has arguments within arguments in different levels and interconnections between the levels. We can extend the Bench-Capon model by using our technique of self-fibring of networks [60]–[35]. This method allows for the recursive substitution of networks inside nodes of other networks [5, 6, 7, 8, 12, 13]. We will work out the details in a later section. Still, we think using LDS is a much better option.

In LDS, this situation will arise if we have a labelled database which includes items such as $t : a, s : b$ and $r : c$ and some additional data, say $u_i : X_i$, such that the following can be proved, among others:³²

- $\gamma(t) : \neg c$
- $\beta(r) : \neg b$
- $\alpha(s) : \neg a$.

α, β, γ are the labels of $\neg a, \neg b$ and $\neg c$ respectively and t, r, s are mentioned in the respective labels to indicate that e.g. $t : a$ is used in the proof of $\gamma(t) : \neg c$ (a with label t attacks c , by proving $\neg c$ with label $\gamma(t)$). The label $\gamma(t)$ shows exactly what role a plays in this attack.

The flattening process acts here as value judgement of what can win, $r : c$ or $\gamma(t) : \neg c$, by comparing r and $\gamma(t)$.

Obviously the value based argumentation machinery can be utilised as part of our flattening mechanism.

The following LDS model will reflect the Bench-Capon coloured diagram:

red: b
blue: a
blue: c
red to blue: $b \rightarrow \neg a$
blue to blue: $a \rightarrow \neg c$
blue to red: $c \rightarrow \neg b$

³²Note that we are assuming here that to defeat x we must put forward an argument for $\neg x$. This is only a simplifying assumption. In LDS, x comes with a label t and so to weaken $t : x$ we can attack t .

Using modus ponens in the form

$$\frac{\alpha : X, \beta : X \rightarrow Y, \varphi(\beta, \alpha)}{\alpha \cup \beta : Y}$$

We can prove:

- red: $\neg a$ if red to blue is allowed
- blue: $\neg c$ if blue to red is allowed
- blue: $\neg b$ if blue to blue is allowed.

The flattening function has to flatten:

- {red: b , blue: $\neg b$ }
- {blue: a , red: $\neg a$ }
- {blue: c , (blue: $\neg c$ is not allowed!)}

Case 1.

red stronger than blue i.e., *blue to red* not allowed.

We get b and $\neg a$ and c .

Case 2.

Blue stronger than red (*red to blue* not allowed)

We get

- {blue: a , (red: $\neg a$ not allowed)}
- {blue: c , blue: $\neg c$ }
- {red: b , blue: $\neg b$ if c is available}

We cannot decide between c and $\neg c$ since both are blue. If we leave them both out or take $\neg c$ then $\neg b$ will not be obtainable and hence we will have $\{a, b\}$.

We see that in the labelled formulation we have more options

1. We can have $X, \neg X$ or neither as choices
2. The label colour (value) can itself be a whole database and so arguments about the values and their strengths can also be part of the system.

The Bench-Capon system is only one level.

The following Figure 5 shows the abstract argumentation structure of Lord Reid's arguments.

Accordingly, Δ_1 in Figure 1 can be better rewritten as Figure 6 below

Assuming that the attack of Lord Reid is successful, then Figure 6 reduces to $\{r_1 : E, r_4 : E \text{ and } s_0 : \neg E\}$. The Lords indeed decided that s_0 was stronger, but they were uncomfortable about it and decided to recommend new legislation.

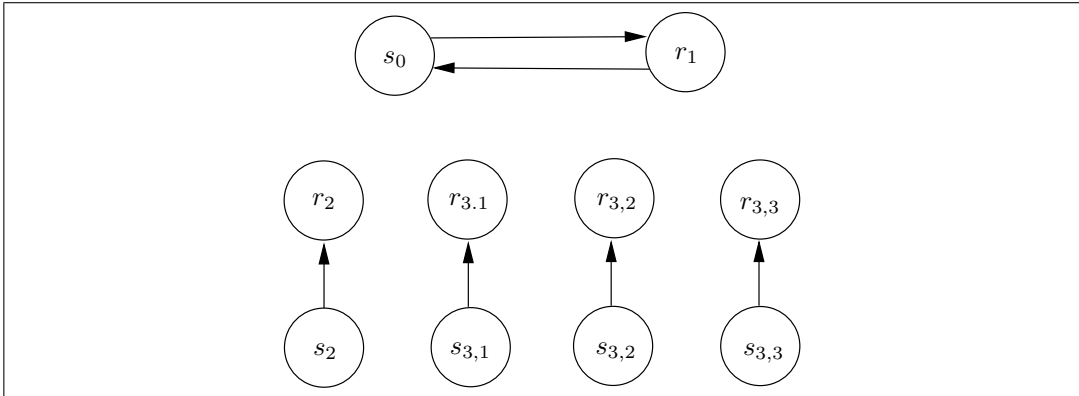


Figure 5

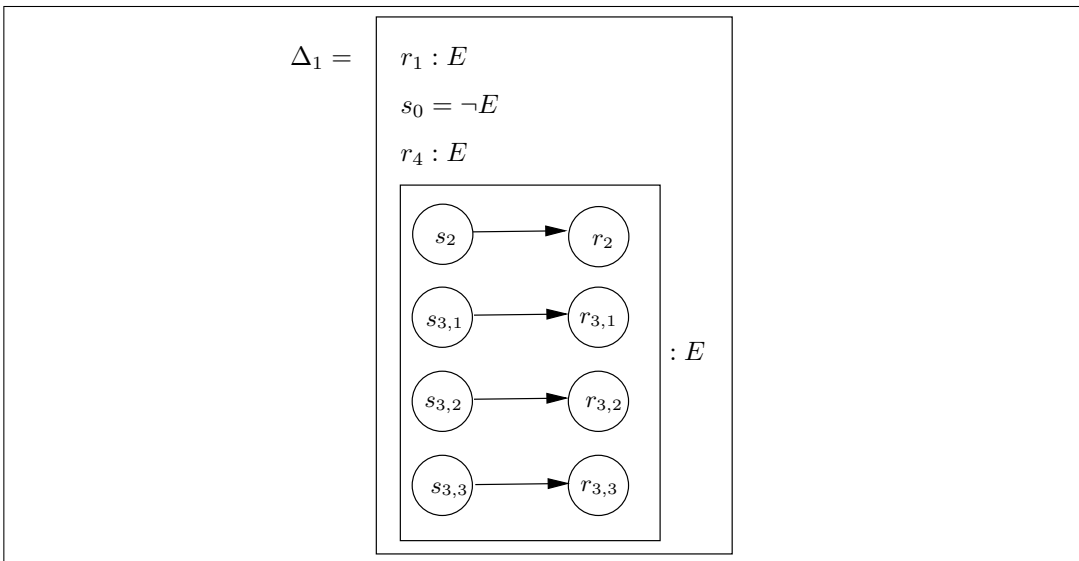


Figure 6

Note Lord Reid’s argument $s_{3,2}$. This is a metalevel value argument like “you cannot colour something red”.

Also note that s_0 and s_2 can be further counter-argued if possible by other Lords. The formal labelling of these additional arguments may require self-fibring. See section 6.5.

6.2 Moral Debate Example

This section also follows Bench-Capon [9, p. 442]. We consider an example cited by Bench-Capon, attributed to Coleman in [22] and Christie [21].

“Hal, a diabetic, loses his insulin in an accident through no fault of his own. Before collapsing into a coma, he rushes to the house of Carla, another diabetic. She is not at home but Hal enters her house and uses some of her insulin. Was Hal justified, and does Carla have a right to compensation?”

The following are the arguments involved as presented in the Bench-Capon paper:

- A = Hal is justified, since a person has a privilege to use the property of others to save their life - the case of necessity.
- B = It is wrong to infringe the property rights of another.
- C = Hal compensates Carla.

Bench-Capon [9] quotes that Christie [21] adds:

- D₁ = If Hal is too poor to compensate Carla, he should nonetheless be allowed to take the insulin, as no one should die because they are poor.
- D₂ = Moreover, since Hal would not pay compensation if too poor, neither should he be obliged to do so even if he can.³³

Bench-Capon further suggests:

- E = Poverty is no defence for theft.
- F = Hal is endangering Carla’s life.
- G = Fact: Carla has abundant insulin.
- H = Fact: Carla does not have ample insulin.

Figure 7 now represents the situation. Note that $H = \neg G$.

Bench-Capon gives the following value properties to the arguments:

³³Christie puts $D_1 + D_2 = D$ together as D. The division into D_1 and D_2 is ours, for later discussion.

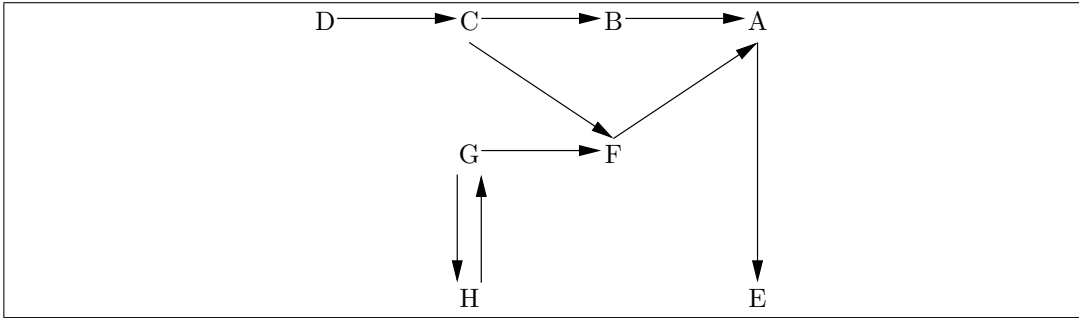


Figure 7

Life: A, D, F
 Property: B, C, E
 Fact: G, H

He says one might argue whether life is stronger than property or not but facts are always the strongest.

Since $H = \neg G$, and since we cannot have both facts, he regards that part of Figure 7 as a case of uncertainty.

We cite this example because we want to analyse what is needed for a better representation of it.

We begin by listing the points:

1. The model needed for a proper analysis of this kind of problem in general (though maybe not necessarily the Hal problem) is a time/action model. There is a difference of values depending at what stage of the action sequence we are at. Has Hal entered Carla's house? Has he checked for insulin? Is it all over and Carla is dead? Each of these cases may have a different argument diagram, possibly with values depending on the previous one! We might add at this point that the need for time/action models has already been strongly emphasised in Gabbay [36] in connection with puzzles involved in the logical analysis of conditionals. This is factors of connected to contrary-to-duty models³⁴ and also needed to incorporate uncertainty. We can get a quite complicated (but highly intuitive) model.³⁵

³⁴See the authoritative survey of A. Jones and J. Carmo [16] in the *Handbook of Philosophical Logic*, 2nd edition.

³⁵We take this opportunity to reinforce our methodological remark of footnote 31. In modelling human practical reasoning, actions and general behaviour it is often a disadvantage and a deficiency to try and use a stylised model and abstract too much from the actual reality (in contrast possibly with modelling physical nature). Often the details of the reality to be modelled suggests the solution

2. We require a better metalevel hierarchy of values and rules, as are available in Labelled Deduction. Possibly such options can also be made adequately available to the abstract argumentation model via self-fibring.
3. The links ($X \rightarrow Y$) should be given strength labels to help us model more realistic cases where an argument X is attacked by arguments Y_1, \dots, Y_k with strength measuring m_1, \dots, m_k .

This is an essential generalisation. One of the quotes we cited from the car case study was (see footnote 18) had the Lords rejecting the written evidence because there was other ample evidence to the same effect (and they didn't want to create a precedent by admitting it).³⁶

4. We can read the link $X \rightarrow Y$ as preventative action of X to stop Y and thus by giving probability of success turn any acyclic network into a Bayesian one. This will introduce uncertainty into the framework. Actually the probability of success is inversely proportional to the conditional probability of Y on X .

6.3 Bayesian aspects of the moral debate example

We begin this section with a closer look at Figure 7. We require a time/action model and contrary-to-duty considerations. We shall explain these features as we model the example.

We imagine an agent, such as Hal, who has available a stock of optional actions. These actions have the form $\mathbf{a} = (A, (B^+, B^-))$ where A is the precondition of the action and B^+, B^- are the post-conditions. A must hold in order for Hal to be allowed to perform the action, in which case the resulting state is guaranteed to satisfy B^+ . However, the agent may take the action anyway, without permission (i.e., A does not hold), in which case the post-condition is B^- . Note that in most cases $B^- = B^+$.

to what otherwise is a puzzle. Let us look at the story and focus on the part which assumes Hal is too poor to replace Carla's insulin. We can ask how is he getting his insulin? Is he getting it on National Health Service? If yes, can't he call the NHS and try to get a replacement? So surely the question of replacement is not 'whether' but 'when', i.e., can he get a replacement in time before Carla runs out of insulin? If life is more important than property this is a good question. If property is more important, then we know he can replace it! Another question, if Hal steals the insulin from Carla and then calls for a replacement, would it not be more difficult to get a replacement (as opposed to calling the NHS first)? We need more details. We are *not* transforming the problem to one more suited to our framework. There are many other examples in other areas which need more details.

³⁶This is a mixture of metalevel/strength/proof argument that only LDS can model. We shall address this kind of argument later.

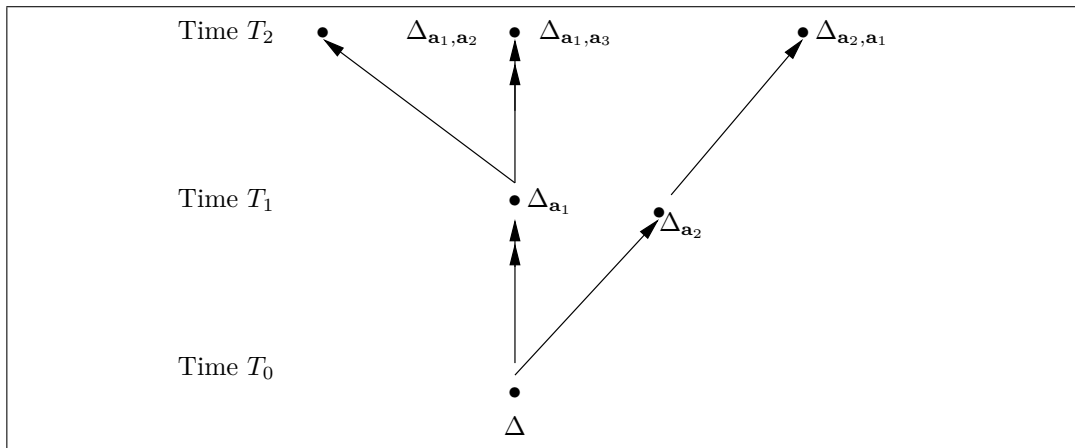


Figure 8

We imagine we are at a state (or time) T_0 , described by a logical theory Δ . The actions allowable to us to perform are $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_i = (A_i, (B_i^+, B_i^-)), \dots$. If $\Delta \vdash A_i$, then action \mathbf{a}_i is allowable at time (state) T_0 , otherwise not. If we perform the action \mathbf{a} , with post-condition B (B is either B^+ or B^-) then we move to time T_1 , with state $\Delta_{\mathbf{a}} = \Delta \circ B$ where $\Delta \circ B$ is the revision of Δ by B . We have $\Delta \circ B \vdash B$.

So to have time action model we need

1. A language for the theories Δ to describe states
2. A language for pre-condition and a language for post-conditions for actions
3. A logic or algorithm for determining when $\Delta \vdash A$ holds, where A is a pre-condition.
4. A revision algorithm giving for each Δ and post-condition B a new theory $\Delta' = \Delta \circ B$. This algorithm can satisfy some reasonable axioms.

Note that the languages for Δ , the pre-conditions and the post-conditions need not be the same!

The flow of time is future branching and is generated by the actions. So if for example our agent can perform actions $\mathbf{a}_1, \dots, \mathbf{a}_k$ as options then after two steps in which he performs say \mathbf{a}_1 first and then say \mathbf{a}_3 , we may get a situation as in Figure 8

The real history at time T_2 is $(\Delta, \Delta_{\mathbf{a}_1}, \Delta_{\mathbf{a}_1, \mathbf{a}_3})$. The states $\Delta_{\mathbf{a}_1, \mathbf{a}_2}$ and $(\Delta_{\mathbf{a}_2}, \Delta_{\mathbf{a}_2, \mathbf{a}_1})$ are hypotheticals.

At time T_0 , our agent chose to take action \mathbf{a}_1 moving onto state $\Delta_{\mathbf{a}_1}$, but he could have chosen to take action \mathbf{a}_2 and done action \mathbf{a}_1 afterwards, ending up at state $\Delta_{\mathbf{a}_2, \mathbf{a}_1}$ at time T_2 . In reality, however, he chose to perform \mathbf{a}_1 and then \mathbf{a}_3 .

The pre-conditions of actions can talk about states and hypotheticals. They need not be in the same language as Δ or the same language as the post-conditions. What is important are the algorithms for ‘ \vdash ’ and ‘ \circ ’.

We are now ready to analyse the moral debate example. First we tell the story in a more realistic way (see footnote 35!). Then we propose some probabilities as an example and we conclude by translating the Bench-Capon statements A–H (page 50) into our time/action set up.

Our story goes as follows. Hal needs insulin. So does Carla. Both are poor and get their insulin from the Health Service. They get it in batches, though not at the same time. So the question whether Carla has spare insulin (G) depends on the time, and is a matter of probability.

Hal loses all his insulin and would need to break into Carla’s property to get hers. He has the option of calling the NHS and asking for replacement, which he can use either for himself if it arrives immediately or to replace Carla’s if necessary. He might get some money from friends. One thing is clear to him. If he steals Carla’s insulin, it will complicate matters; it might be more difficult to find a replacement. So the question of compensation C is also a matter of probability. The following are the possible scenarios.

If property is valued more than life, then if Hal steals Carla’s insulin, the probability of getting a replacement is lower in the case where Carla’s life is not threatened.

If life is valued more than property, his chances of obtaining replacement is higher in case Carla’s life is threatened.

We must clarify what ‘getting a replacement’ means. Hal will probably start a process for getting insulin for himself immediately at start time T_0 . Since it might not arrive in time, he will break into Carla’s home and use hers, and hope to use the insulin he ‘ordered’ to replace Carla’s. If Carla has ample insulin, there is a higher chance or that the replacement will arrive in time before Carla’s life is threatened. If Carla does not have ample insulin, Hal can use this as a further reason to rush the process of replacement. This further reason might be counterproductive if property is valued above life.

So the statement

C = Hal gets a replacement
should be taken as (see footnote 35):

Hal gets a replacement before Carla is in need of it.

We may then have the following scenarios (P stands for Probability $P(x)$ and it should be indexed by case and time, i.e., $P_{1,a}, P_{1,b}, P_{2,a}$ and $P_{2,b}$:

Case 1. Property stronger than life

- (a) Time = Before Hal breaks into Carla's house.

$$P(G) = \frac{2}{3}$$

$$P(\neg G) = \frac{1}{3}$$

$$P(C/G) = 0.9$$

$$P(\neg C/G) = 0.1$$

(Since Carla does have ample insulin, Hal has more time to replace what he might take.)

$$P(C/\neg G) = 0.5$$

$$P(\neg C/\neg G) = 0.5$$

(Admittedly, Carla's life is in danger but there may not be enough time to get a replacement. On the other hand, this very fact might help get the insulin more quickly. Note that the event C means 'getting replacement in time'.)

- (b) Time = After Hal breaks into Carla's house.

At this stage the value of G is known: either $G = 1$ or $G = 0$. We get

$$P(C/G = 1) = 0.7$$

$$P(\neg C/G = 1) = 0.3$$

(less than before breaking into the house, because Hal committed a serious crime. He may not be favourable with the authority.)

$$P(C/G = 0) = 0,4$$

$$P(\neg C/G = 0) = 0.6$$

Again, less than before. See also Gabbay and Woods [56].

Case 2. Property not stronger than life³⁷

- (a) Time = Before Hal breaks into Carla's house

$$P(G) = \frac{2}{3}$$

$$P(\neg G) = \frac{1}{3}$$

$$P(C/G) = 0.9$$

$$P(\neg C/G) = 0.1$$

$$P(C/\neg G) = 0.9$$

$$P(\neg C/\neg G) = 0.1$$

³⁷Jon Williamson reminded us that it is reasonable to assume that the legal process does not make general value judgements like this, nor can a legal argument appeal to such judgements. Instead much more specific 'mitigating circumstances' can be used to reduce the length of a sentence on conviction ('I did it to save my life, guv').

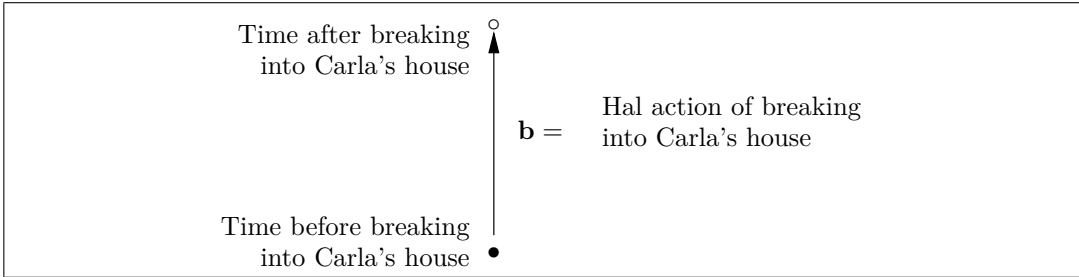


Figure 9

(b) Time = After Hal breaks into Carla's house

$$P(G) = \frac{2}{3}$$

$$P(\neg G) = \frac{1}{3}$$

$$P(C/G = 1) = 0.9$$

$$P(\neg C/G = 1) = 0.1$$

$$P(C/G = 0) = 0.7$$

$$P(\neg C/G = 0) = 0.3.$$

Let us now translate the arguments involved in the original moral debate example of Section 6.2.

When is Hal justified in breaking into Carla's home? The answer is yes only in the case that life is stronger than property and he can reasonably say he is not risking her life. That depends on finding a replacement. We therefore have to calculate the probability of C given all the data we have.

Thus our time/action axis has the form of Figure 9:

The actions available to Hal are:

1. \mathbf{b} = breaking into Carla's house. The post-condition is breaking in and taking the insulin. The pre-condition of \mathbf{b} is high probability of replacing Carla's insulin (in time before she needs it) in case *life is stronger than property* and \perp (falsity i.e., no permission to do the action) in case *life is not stronger than property*.
2. \mathbf{r} = actions having to do with getting a replacement of insulin. We assume he can perform these actions at any time but the post-conditions are not clear.³⁸

We need also agree the value of the threshold probability, e.g. only if there is at least 0.9 chance of replacement can Hal break into Carla's home to take the insulin. Consider now:

³⁸We may need a temporal language for the post-conditions so that we can say something like 'insulin will be delivered in two days'.

B = It is wrong to infringe the property of others.

B is an argument reflected in the pre-condition of the action \mathbf{b} , it can be done when B satisfied otherwise not. I would write it as

$$\mathbf{b} = (\text{Justification, Break in and taking insulin}).$$

Let us now model the chain of events as a Bayesian network. The story is clear. Depending on the probability $P(G)$, Hal decides whether he wants to break into Carla's house \mathbf{b} (no use breaking into her house if she does not have enough insulin). He is justified J in breaking \mathbf{b} into Carla's house if there is high probability of compensation C . Thus C depends both on \mathbf{b} and G , and \mathbf{b} also depends on G . We have the following network, Figure 10.

There are two problems with this representation.

1. The dependency of \mathbf{b} on G is not on $G = 1$ or $G = 0$ but on $P(G)$. Say if $P(G) < 0.1$ then maybe $\mathbf{b} = 0$.

This is OK because the probabilities can be made to take account of that. This is allowed in the theory of Bayesian nets.

2. The probabilities in Figure 10 depend on whether property is stronger than life or not. The best way to represent this is to have a Bayesian net with one variable only, *Case*.

Case =1 means property stronger than life and *case* =0 means property is not stronger than life.

For each case we get a different copy of Figure 10 with different probabilities.

So we get a substitution of the network of Figure 10 into a one point network:

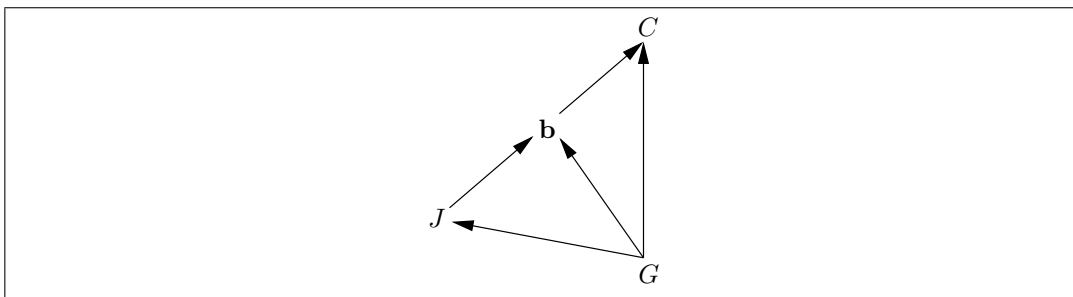


Figure 10

- *Case*. This operation is in accordance with the ideas in [113].

We can also allow for several justification variables to make it more realistic.

It is not difficult to work out the details of the rest of C–H, but the reader can already see that in the simple minded model there is lack of sensitivity to a variety of metalevels.

6.4 Neural Representation of Argumentation Frameworks

This subsection, based on [61] will outline how to represent (in neural nets) any value-based argumentation framework involving x and anti- x (i.e., arguments and counter-arguments). For instance, it can be implemented in neural networks with the use of Neural-Symbolic Learning Systems [58]. A neural network consists of interconnected neurons (or processing units) that compute a simple function according to the weights (real numbers) associated to the connections. Learning in this setting is the incremental adaptation of the weights [68]. The interesting characteristics of neural networks do not arise from the functionality of each neuron, but from their collective behaviour, thus being able to efficiently represent (and learn) multi-part, cumulative argumentation, as exemplified below.

Cumulative behaviour can be encoded in Neural-Symbolic Learning Systems with the use of a hidden layer of neurons in addition to an input and an output layer in a feedforward network. Rules of the form $A \wedge B \rightarrow C$ can be represented by connecting input neurons that represent concepts A and B to a hidden neuron, say h_1 , and then connecting h_1 to an output neuron that represents C in such a way that output neuron C is activated (true) if input neurons A and B are both activated (true). If, in addition, a rule $B \rightarrow C$ is also to be represented, another hidden neuron h_2 can be added to the network to connect input neuron B to output neuron C in such a way that C is now activated also if B alone is activated.³⁹ This is illustrated in Figure 11. The network can be used to perform the computation of the rules in parallel such that C is true whenever B is true [58].

In a neural network, positive weights can represent the support for an argument, while negative weights can be seen as an attack on an argument. Hence, a negative weight from a neuron A to a neuron B can be used to implement the fact that A attacks B . Similarly, a positive weight from B to itself can be used to indicate that B supports itself. Since we concentrate on feedforward networks, neuron B will

³⁹In the general case, hidden neurons are necessary to implement the following conditions: **(C1)** The input potential of a hidden neuron (N_i) can only exceed N_i 's threshold (θ_i), activating N_i , when all the positive antecedents of r_i are assigned the truth-value *true* while all the negative antecedents of r_i are assigned *false*; and **(C2)** The input potential of an output neuron (A) can only exceed A 's threshold (θ_A), activating A , when at least one hidden neuron N_i that is connected to A is activated.

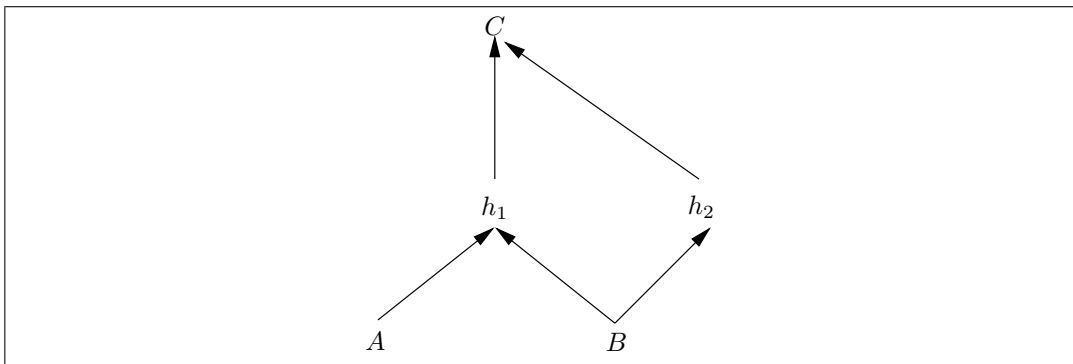


Figure 11: A simple example of the use of hidden neurons

appear on both the input and the output layers of this network as shown in Figure 12, in which dotted lines are used to indicate negative weights.

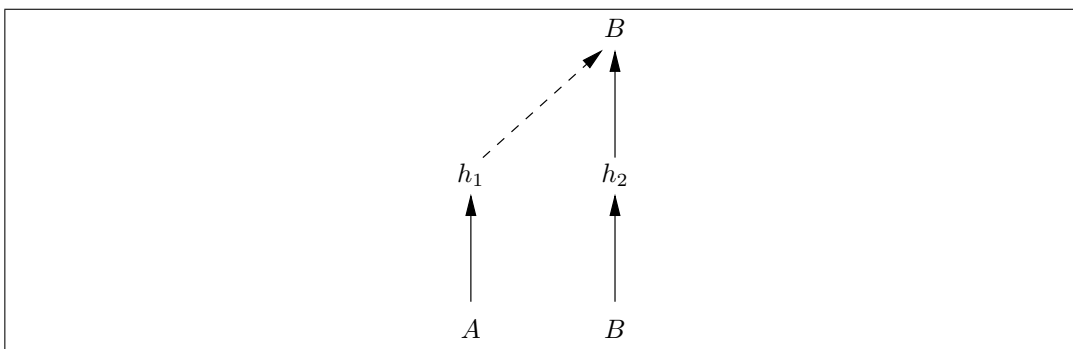


Figure 12: A simple example of the use of negative weights for counter-argumentation

In Figure 12, *A* attacks *B* via h_1 , while *B* supports itself via h_2 . Suppose now that, in addition, *B* attacks *C*. We need to connect input neuron *B* to output neuron *C* via a new hidden neuron h_3 . Since *B* appears on both the network’s input and output, we also need to add a feedback connection from output neuron *B* to input neuron *B* such that the activation of *B* can be computed by the network according to the chain ‘*A* attacks *B*’, ‘*B* attacks *C*’, etc. As a result, in Figure 13 (in which we do not represent *B*’s feedback connection for the sake of clarity), if the attack from *A* on *B* is stronger (according to the network’s weights) than *B*’s support to itself, then *A* will block the activation of (output) *B*, and (input) *B* will not be able to block the activation of *C*. In this case, the network’s final computation will include *C* and not *B* in a stable state. If, on the other hand, *A* is not strong enough to

block B , then B will be activated and block C .

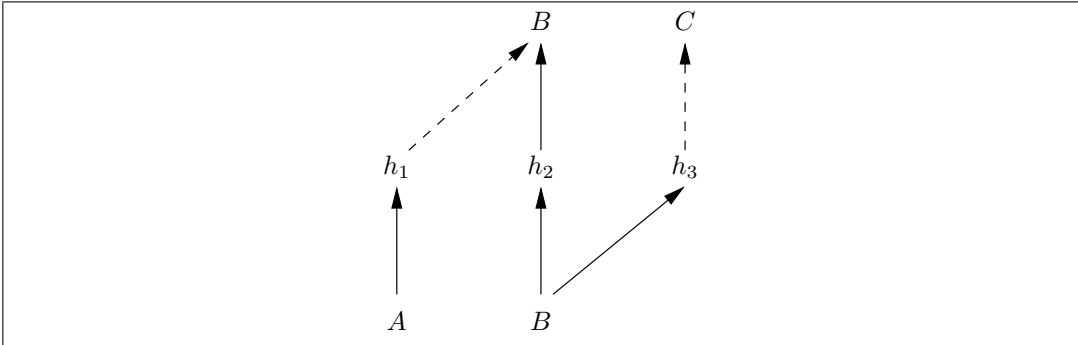


Figure 13: The computation of arguments and counter-arguments

Let us take the example in which an argument A attacks an argument B , and B attacks an argument C , which in turn attacks A in a cycle. In order to implement this in a neural network, we need positive weights to explicitly represent the fact that A supports itself, B supports itself and so does C . In addition, we need negative weights from A to B , from B to C and from C to A (see Figure 14) to implement attacks. If all the weights are the same in absolute terms, no argument wins, as one would expect, and the network stabilises with none of $\{A, B, C\}$ activated. If, however, the value of A (i.e., the weight from h_1 to A) is stronger than the value of C (the weight from h_3 to C , which is expected to be the same in absolute terms as the weight from h_3 to A), C cannot attack and defeat A . As a result, A is activated. Since A and B have the same value (as e.g., in the previous case of an unspecified priority), B is not activated, since the weights from h_1 and h_2 to B will both have the same absolute value. Finally, if B is not activated then C will be activated, and a stable state $\{A, C\}$ will be reached in the network. In Bench-Capon's model [9], this is exactly the case in which colour blue is assigned to A and B , and colour red is assigned to C with blue being stronger than red. Note that the order in which we reason does not affect the final result (the stable state reached). For example, if we started from B successfully attacking C , C would not be able to attack A , but then A would successfully attack B , which would this time round not be able to successfully attack C , which in turn would be activated in the final stable state $\{A, C\}$. This indicates that a neural (parallel) implementation of this reasoning process could be advantageous also from a purely computational point of view.

Note that (as in the general case of argumentation networks) in the case of neural networks, we can extend Bench-Capon's model with the use of self-fibring neural networks, which allow for the recursive substitution of neural networks inside

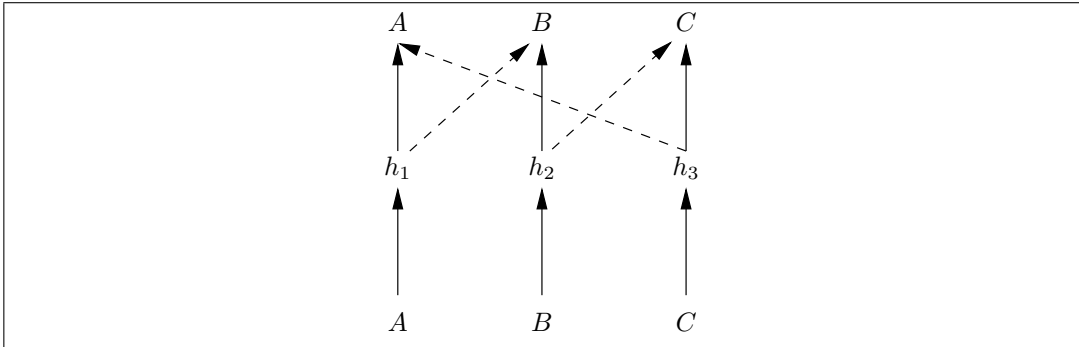


Figure 14: The moral-debate example as a neural network

nodes of other networks [60].

The implementation of the network’s behaviour (weights and biases) must be such that, when we start form a number of positive arguments (input vector $\{1, 1, \dots, 1\}$), weights with the same absolute values cancel each other producing zero as the output neuron’s input potential. A neuron with zero or less input potential is then deactivated, while a neuron with positive input potential is activated. This allows for the implementation of the argumentation framework in neural-symbolic learning systems, in the style of the translation algorithms developed at [59].

6.5 Self-fibring of Argumentation Networks

We will conclude this section by indicating how to do self-fibring of argument networks. The mechanics of it is simple. We begin with one network, say the one in Figure 4. We pick a node in it, say node a , and substitute another network for that node, say we substitute the network of Figure 7. We thus get the ‘network’ of Figure 15.

The need of self-fibring may arise if additional arguments are available supporting the contents of the node.

The self-fibring problem has three aspects:

Aspect 1: Intuitive Meaning

What is the intended interpretation/meaning of this substitution? This can be decided by the needs of the application area. Here are some options:

- (1.1) a is supposed to be an argument, so Figure 7 can be viewed as delivering some winning argument (A of Figure 7) which can combine/support a .
- (1.2) Figure 7 is a network so b of Figure 4 can plug into it. We can connect b to all

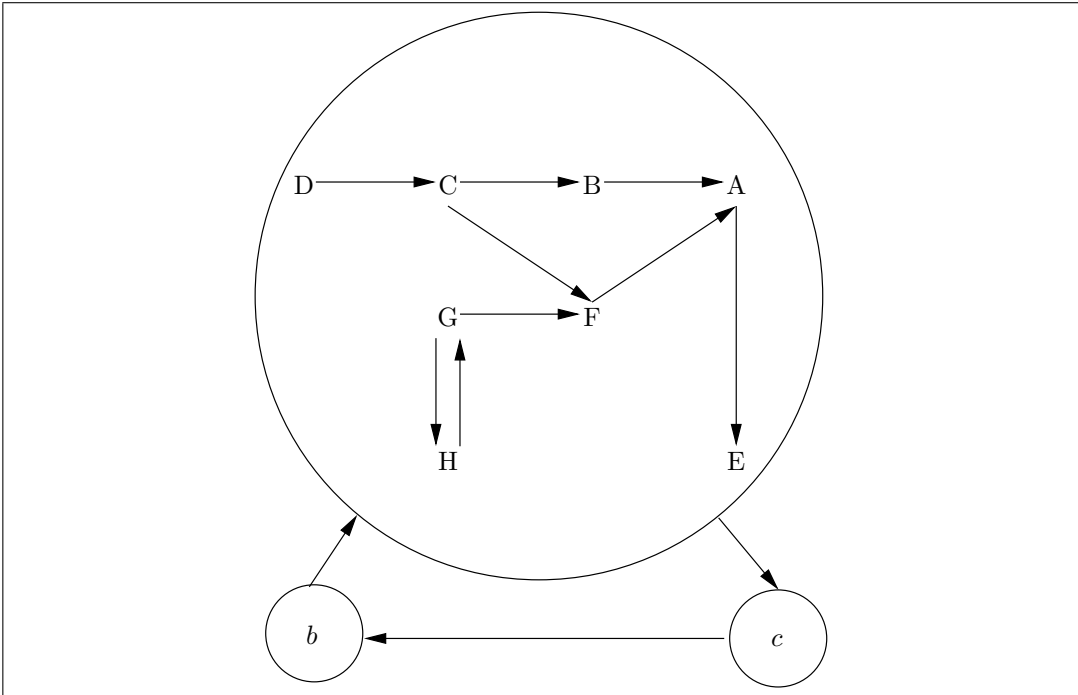


Figure 15

(or some) members of Figure 4 and similarly connect all (or some) members of Figure 7 into c of Figure 4.

For various options see [113, 60, 35].

Aspect 2: Formal aspect

(2.1) *Syntactical substitution*

Formally the node a is supposed to be an argument. So we need a fibring function $\mathbf{F}(\text{node, network}) = e$ yielding a node e and so we end up with Figure 16

\mathbf{F} might do, for example, the following: \mathbf{F} can use the colour of node a to modify the colours of the nodes in Figure 7 (the substituted network), and maybe also modify some connections in Figure 7, and then somehow emerge with some winning argument e and a colour to be substituted/combined with a and its colour.

(2.2) *Semantic substitution*

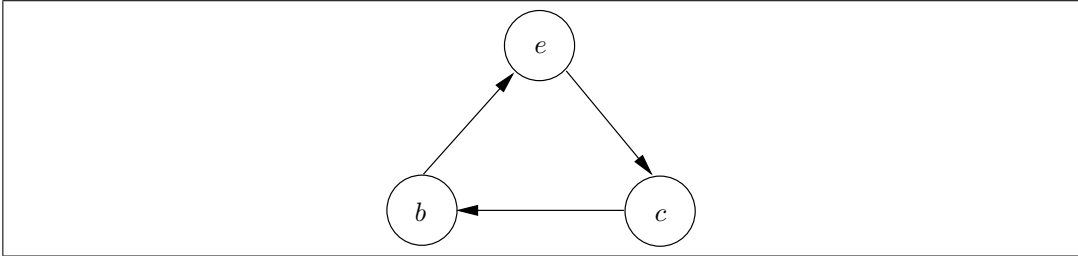


Figure 16

If the original network has an interpretation, then the node a can get several possible semantic values. We can make the definition of the substitution context sensitive to those values. We may even go to the extent of substituting different networks for different options of values.

Aspect 3: Coherence

To enable successful repeated recursive substitution of networks within networks, we have to modify our definition of the original network. For example:

- (3.1) Possibly extend the notion of network and allow arrows to either support or defeat arguments.
- (3.2) Restrict the substitution of networks for nodes by compatibility/consistency conditions.

Example: Self-fibred argumentation network

We have a set of nodes and links of the form (a, b) meaning a attacks b . We also have valuation colours. A weaker colour cannot attack a stronger colour. So far this is the Bench-Capon definition.

Let a be a node. Define the notion of x is a supportive (resp. attacking) node for a as follows:

- a is supportive of a
- if x is supportive (resp. attacking) node of a and y attacks x then y is an attacking (resp. supportive) node of a .

Now let a be a node in a network A and suppose we have another network N which we want to substitute for a . We must assume a appears in N with the same colour value as it is in A . We substitute N for a and make new connection as follows:

- Any node x of A which attacks a in A is now connected to any node y in N which supports a in N .

- Any node y in N which supports a in N is now made connected to any node x of A which a of A is attacking.

This definition is reasonable. a is an argument in network A . N is another network which is supposed to support a (a in N). Thus anything which attacks a in A will attack all a supporters in N and these in turn will attack whatever nodes a attacks in A . Note that he may be attacking facts in N by this wholesale connection of arrows. However, Bench-Capon has already remarked that facts should get the strongest colour and so the colours will take care of that!

See reference [42].

Acknowledgements

We are grateful to A. Garcez, L. Lamb, D. Makinson, G. Pigozzi and J. Williamson for valuable comments.

References

- [1] Ruggero J. Aldisert. *Logic for Lawyers: A Guide to Clear Legal Thinking*, Clark Boardman, 1989.
- [2] A. Aliseda. *Seeking Explanation. Abduction in Logic, Philosophy of Science and Artificial Intelligence*, ILLC, 1999.
- [3] Christopher Allen. *Practical Guide to Evidence*, 2nd ed. Cavendish Pub, 2001.
- [4] H. Barringer, D. M. Gabbay, and J. Woods. Temporal Dynamics of Argumentation Networks. In Volume Dedicated to Joerg Siekmann, D. Hutter and W. Stephan, editors. *Mechanising Mathematical Reasoning*, Springer Lecture Notes in Computer Science 2605, pp. 59-98, 2005.
- [5] H. Barringer, D. M. Gabbay and J. Woods. Temporal dynamics of support and attack networks: From argumentation to Zoology. In Hutler D and Stephan W (eds). *Mechanizing Mathematical Reasoning: Essays in Honor of Jörg Siekman on the Occasion of his 60th Birthday*, pages 59-98, Springer-Verlag, 2005.
- [6] H. Barringer, D. M. Gabbay and J. Woods. Network modalities. In Gross G and Schulz U, eds. *Linguistics, Computer Science and Language Processing : Festschrift for Franz Guentner on the Occasion of his 60th Birthday*, College Publications, 2008.
- [7] H. Barringer, D. M. Gabbay and J. Woods. Temporal argumentation networks. *Argumentation and Computation*, 2-3, 143-202, 2012.
- [8] H. Barringer, D. M. Gabbay and J. Woods. Modal argumentation networks. *Argument and Computation*, 2-3, 203-227, 2012.

- [9] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, **13**, 429–448, 2003.
See also: <http://www.csc.liv.ac.uk/~tbc/FTP/kings2.ppt>
- [10] T. Bench-Capon. Agreeing to differ: modelling persuasive dialogue between parties without a consensus about values. *Informal Logic*, forthcoming, 2003.
- [11] T. Bench-Capon. Knowledge based systems in the legal domain. A survey article in *Encyclopaedia of Computer Science and Technology*, available on the web at <http://www.csc.liv.ac.uk/~lial/lial/tut.html>
- [12] J. F. Benthem. *Logical Dynamics of Information and Interaction*, Cambridge University Press, 2011.
- [13] J. F. Benthem. The nets of reason. *Argument and Computation*, 2-3, 83-86, 2012.
- [14] G. Bongiovanni, G. Postema, A. Rotolo, G. Sartor, and D. Walton (eds). *Handbook of Legal Reasoning and Argumentation*, Cambridge Univ Press, 2016.
- [15] P. D. Bruza, D. Widdows and J. Woods. A quantum logic for down below. In Engesser K, Gabbay DM and Lehmann D (eds). *Handbook of Quantum Logic and Quantum Structures: Quantum Logic*, 625-660, North-Holland, 2009.
- [16] J. Carmo and A. Jones. Deontic logic and contrary-to-duties. *Handbook of Philosophical Logic*, Volume 8, 2nd edition, pp. 265–345, D. M. Gabbay and F. Guenther, eds. Kluwer, 2002.
- [17] W. Carnielli, M. Coniglio, D. M. Gabbay, P. Gouveia and C. Sernadas. *Analysis and Synthesis of Logics*. Springer, 2007, 500p.
- [18] W. Carnielli, M. E. Coniglio and I. M. Loffredo D’Ottaviano, (eds). *The Many Sides of Logic*, College Publications, 2009.
- [19] C. I. Chesnevar, A. G. Maguitman and R. P. Loui. Logical models of argument. *ACM Computing Surveys (CSUR) Surveys Homepage archive* Volume 32 Issue 4, Dec. 2000 Pages 337-383.
- [20] D. Chiffi and A. V. Pietarinen. The extended Gabbay-Woods schema and scientific practices. In Gabbay *et al.*, 2019. 331-348, 2019.
- [21] G. C. Christie. *The Notion of an Ideal Audience in Legal Argument*. Kluwer, 2000.
- [22] J. Coleman. *Risks and Wrongs*. Cambridge University Press, 1992.
- [23] I. Copi. *Introduction to Logic*, Macmillan, 1953.
- [24] Sir Rupert Cross. *On Evidence*. Butterworth, 1999.
- [25] M. L. Dalla Chiara, R. Guintini and M. Rédei, (eds). The history of quantum logics. In Gabbay DM and Woods J (eds). *The Many Valued and Nonmonotonic Turn in Logic*, volume 8 of Gabbay and Woods (eds.) *Handbook of the History of Logic*, 205-284, North-Holland, 2007.
- [26] H. Dennis. *Law of Evidence*, Sweet and Maxwell, 1992.
- [27] Deon Conferences. Journals.
Please see: <http://www.doc.ic.ac.uk/deon02/>, <http://www.cert.fr/deon00/>

<http://www.denniskennedy.com/ailaw.htm>, <http://www.iaail.org/>.

There are many useful Links to Evidence-Related Web Sites: at <http://tillers.net>

- [28] K. Engesser, D. M. Gabbay, and D. Lehmann. Nonmonotonicity and holicity in quantum logic. In Engesser *et al.*, 587-624, 2009.
- [29] G. Englebretsen. Is natural logic part of naturalized logic? In Gabbay *et al.*, 593-620, 2019.
- [30] M. Fisher, D. M. Gabbay, and L. Vila, (eds). *Handbook of Temporal Reasoning in Artificial Intelligence: Foundations of Artificial Intelligence 1*, Elsevier, 2005.
- [31] P. A. Flack and A. Kakas, (eds). *Abduction and Induction: Essays on Their Relation and Integration*, Kluwer, 2000.
- [32] D. M. Gabbay. *Handbook of Logic and Computer Science, 1 Background: Mathematical Structures, and 2 Background: Computational Structures*, Clarendon Press, 1992.
- [33] D. M. Gabbay. Logic made reasonable. *Kunstliche Intelligenz*, 6, 39-41, 1992.
- [34] D. M. Gabbay. *Labelled Deduction Systems*. Oxford University Press, 1996.
- [35] D. M. Gabbay. *Fibiring Logics*, Oxford University Press, 1998.
- [36] D. M. Gabbay. Dynamics of practical reasoning, a position paper. In *Advances in Modal Logic 2, Proceedings of Conference October 1999*, K. Segerberg *et al.*, eds, pp. 179–224. CSLI Publications, Cambridge University Press, 2001.
- [37] D. M. Gabbay. Sampling LDS. In *A Companion to Philosophical Logic*, D. Jacquette, ed., pp. 742–769. Blackwell, 2002.
- [38] D. M. Gabbay. Editorial to *Handbook of Philosophical Logic*, 2nd edition. Kluwer, 2002–2012.
- [39] D. M. Gabbay. Reactive Kripke models and contrary-to-duty obligations. DEON-2008, *Deontic Logic in Computer Science*, Ron van der Meyden and Leendert van der Torre, eds. LNAI 5076, pp. 155–173, Springer, 2008.
- [40] D. M. Gabbay. Semantics for higher level attacks in extended argumentation frames. Part 1: Overview. *Studia Logica*, 93:355–379, 2009.
- [41] D. M. Gabbay. Modal foundations for argumentation networks. *Studia Logica*, 93(2-3): 181–198, 2009.
- [42] D. M. Gabbay. Fibiring argumentation frames. *Studia Logica*, 93(2-3), 231-295, 2009.
- [43] D. M. Gabbay. Reactive Kripke models and contrary-to-duty obligations. Expanded version to appear in *Journal of Applied Logic*.
- [44] D. M. Gabbay and M. Gabbay. Argumentation as information input, 2015. Short version published in *Proceedings COMMA 2016, Computational Models of Argument*, Pages 311 - 318 DOI10.3233/978-1-61499-686-6-311 A Volume in Series Frontiers in Artificial Intelligence and Applications IOS press Volume 287: Full version published by College Publication in the G. Simari Tribute *Argumentation-based Proofs of Endearment, Essays in Honor of Guillermo R. Simari on the Occasion of his 70th Birthday*, Carlos I Chesnevar, Marcelo A Falappa, Eduardo Ferme, eds. pp 145-197
- [45] D. M. Gabbay and A. d’Avila Garcez. Logical modes of attack in argumentation net-

- works. *Studia Logica*, 93(2-3): 199–230, 2009.
- [46] D. M. Gabbay and A. Hunter. Making inconsistency respectable. A logical framework for inconsistency in reasoning. In Jourand Ph and Kelemen T (eds). *Fundamentals of Artificial Intelligence Research*, 19-32, Springer, 1991.
- [47] D. M. Gabbay and A. Hunter. Making inconsistency respectable part 2: Meta-level handling of inconsistency. In Clarke M, Kruse R and Seraffin S (eds). *Lecture Notes on Computer Science*, 129-136, Springer, 1992.
- [48] D. M. Gabbay and G. Rozenberg. Reasoning Schemes, Expert Opinion and Critical Questions: Sex Offenders Case Study, *IFCoLog Journal of Logics and their Applications* Volume 4, Number 6 July 2017, PP 1687-1789.
- [49] D. M. Gabbay and J. Woods. Cooperate with your logic ancestors. *Journal of Logic, Language and Information*, **8**, iii–v, 1999.
- [50] D. M. Gabbay and J. Woods. *Agenda Relevance: A Study in Formal Pragmatics*, Elsevier, 2003, 521 pp.
- [51] D. M. Gabbay and J. Woods. *Agenda Relevance: A Study in Formal Pragmatics*, North-Holland, 2003.
- [52] D. M. Gabbay and J. Woods. *The Reach of Abduction: Insight and Trial*, Elsevier, 2005, 476 pp.
- [53] D. M. Gabbay and J. Woods. *The Reach of Abduction: Insight and Trial*, North-Holland, 2005.
- [54] D. M. Gabbay and J. Woods. Probability in the law. In *Dialogues, Logics and Other Strange Things: Essays in Honour of Shahid Rahman*, Cédric Dégrémont, Laurent Keiff and Helge Rückert, eds. College Publications, 2008.
- [55] D. M. Gabbay, A. S. d’Avila Garcez and L. C. Lamb. *Connectionist Non-classical Logics: Distributed Reasoning and Learning in Neural Networks*. Monograph, Springer Verlag, 2008.
- [56] D. M. Gabbay and J. Woods. Probability in the law. In Dégrémont C, Keiff L and Rukert H, eds. *Dialogues, Logics and Other Strange Things: Essays in Honour of Shahid Rahman*, College Publications, 2008.
- [57] D. M. Gabbay and J. Woods. Logic and the law: Crossing the lines of discipline. In Gabbay DM, Canivez P, Rahman S and Thiercelin (eds). 2010. *Approaches to Legal Rationality*, 165-202. Springer, 2010.
- [58] A. S. d’Avila Garcez, K. Broda and D. M. Gabbay, *Neural-Symbolic Learning Systems*, Springer-Verlag, 2002.
- [59] A. S. d’Avila Garcez, L. C. Lamb, D. M. Gabbay and K. Broda, Connectionist Modal Logics for Distributed Knowledge Representation, submitted, 2002.
- [60] A. S. d’Avila Garcez and D. M. Gabbay, Fibring Neural Networks, In *Proceedings of 19th National Conference on Artificial Intelligence (AAA’04)*, pp. 342-347. San Jose, CA. AAAI Press, 2004.
- [61] A. S. D’Avila Garcez, D. M. Gabbay and L. Lamb. Value based argumentation frameworks as neural networks. *Journal of Logic and Computation*, 15(6):1041-1058, Dec. 2005.

- [62] J. Y. Girard, Y. Lafont and P. Taylor. *Proofs and Types*. Cambridge University Press, 1989.
- [63] T. F. Gordon. *The Pleadings Game: An Artificial Intelligence Model of Procedural Justice*, Kluwe, 1995r.
- [64] T. F. Gordon, H. Prakken and D. Walton. The Carneades model of argument and burden of proof, *Artificial Intelligence*, 171, 875-896, 2007.
- [65] J. W. Guan and D. A. Bell. *Evidence Theory*, 2 volumes. Elsevier, 1991.
- [66] C. L. Hamblin. *Fallacies*, Methuen, 1970.
- [67] J. W. Harris. *Legal Philosophies*, 2nd edition, 2003.
- [68] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd edition, Prentice Hall, 1999.
- [69] V. E. Hendricks. *Mainstream and Formal Epistemology*, Cambridge University Press, 2006.
- [70] C. Hewitt and J. Woods, (eds). *Inconsistency Robustness*, College Publications, 2015.
- [71] J. Hintikka. *Inquiry as Inquiry: A Logic of Scientific Discovery*, Reidel, 1999.
- [72] A. Keane. *The Modern Law of Evidence*, Butterworth, 2000.
- [73] J. Keppens. Conceptions of vagueness in subjective probability for evidential reasoning. In Governatori (ed). *Proceedings of 22nd Annual Conference on Legal Knowledge and Information Systems*, 79-999 IOS Press, 2009.
- [74] S. Kripke. A completeness theorem in modal logic, *Journal of Symbolic Logic*, 24, 1-14, 1959.
- [75] S. Kripke. Semantical considerations on modal logic, *Acta Philosophica Fennica*, 16, 83-94, 1963.
- [76] T. A. F. Kuipers. Abduction aiming at empirical progress of even truth approximation leading to a challenge for computational modelling. *Foundations of Science*, 4, 307-323, 1999.
- [77] E. F. Loftus. Leading questions and eyewitness reports *Cognitive Psychology*, 7, 560-572, 1975.
- [78] E. F. Loftus, D. Wolchover, and D. Page. General review of the psychology of witness testimony, in Heath-Armstrong A. *et al.* (eds.) *Witness Testimony: Psychological, Investigative and Evidential Perspectives*, Oxford Univ Press, 2006.
- [79] R. P. Loui. Process and policy: Resource-bounded non-demonstrative reasoning, *Computational Intelligence* 14 (1998) 138.
- [80] M. MacCrimmon and P. Tillers, eds. *The Dynamics of Judicial Proof*. Physica-Verlag, 2002.
- [81] L. Magnani. *Abduction, Reason and Science: Processes of Discovery and Explanation*, Kluwer, 2001.
- [82] L. Magnani. *Abductive Cognition. The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*. Springer, 2009.
- [83] L. Magnani. Naturalizing logic and errors of reasoning vindicated: Logic reapproaches

- cognitive science. *Journal of Applied Logic*, 13, 13-36. 2015.
- [84] L. Magnani. *The Abductive Structure of Scientific Creativity. An Essay on the Ecology of Cognition*, Springer, 2017.
- [85] L. Magnani. The urgent need of a naturalized logic. In Dodig-Crmkovic G and Schroeder MJ (eds). *Contemporary Natural Philosophy and Philosophies*, a special guest-edited number of *Philosophies*, 3, 44, 2018.
- [86] L. Magnani. Errors of reasoning exculpated: Naturalizing the logic of abduction. In Gabbay *et al.* 2019, 269-308, 2019.
- [87] C. Mortensen. *Inconsistent Mathematics*, Kluwer, 1995.
- [88] C. Mortensen. *Inconsistent Geometry*, College Publications, 2010.
- [89] I. Niiniluoto. *Truth-Seeking by Abduction*, Springer, 2018.
- [90] W. Park. *Abduction in Context. The Conjectural Dynamics of Scientific Reasoning*, Springer, 2017.
- [91] Ch. Perelman. *Justice, Law and Argument*. Reider, 1980.
- [92] H. Prakken. *Logical Tools for Modelling Legal Argument*, Kluwer, 1997.
- [93] H. Prakken and G. Sartor. On modelling burdens and standards of proof in structural argumentation. In Atkinson KD (ed). *Legal Knowledge and Information Systems*, 83-92, IOS Press. 2011.
- [94] H. Prakken, C. Reed and D. Walton. Argumentation schemes and generalisation in reasoning about evidence. *ICAIL-03*, June 24–28, 2003.
- [95] H. Putnam. Is logic empirical? In Cohen R and Warofsky M (eds). *Boston Studies in the Philosophy of Science*, 174-197, Reidel, 1968.
- [96] M. Ransom. Naturalizing logic: A case study of the ad hominem and implicit bias. In Gabbay *et al.*, 573-589, 2019.
- [97] C. Reed and D. Walton. Argument schemes in dialogue, dissensus and the search for common ground. In H. V. Hansen, C. W. Tindale, J. A. Blair, R. H. Johnson and D. M. Godden (eds). *Proceedings of OSSA*, Windsor, ON (CD-ROM), 2007.
- [98] Report of the Royal Commission of Criminal Justice, M 2263, 1993. London: HMSO, Ch 8, para 26.
- [99] M. J. Sergot, R. A. Kowalski *et al.*. British Nationality Act. *Communications of the ACM*, **29**, 370–386, 1986.
- [100] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [101] S. Toulmin. *The Uses of Argument*, Cambridge University Press, 1958.
- [102] S. Uglow. *Textbook on Evidence*, Sweet and Maxwell, 1997. (Second edition: Evidence Text and Materials Paperback — 29 Aug 2006).
- [103] B. Verheij. *Virtual Arguments. On the Design for Lawyers and Other Arguers*, The Asser Press, 2005.
- [104] D. Walton. *Appeal to Expert Opinion*. Penn State University Press, 1997.
- [105] D. Walton. *Legal Argument and Evidence*, Penn State University Press, 2002.
- [106] D. Walton. *Abductive Reasoning*, University of Alabama Press, 2004.

- [107] D. Walton. *Argument Methods for Artificial Intelligence in Law*, Springer, 2005.
- [108] D. Walton. *Character Evidence: An Abductive Theory*, Springer, 2007.
- [109] D. Walton. *Witness Testimony Evidence*, Cambridge University Press, 2008.
- [110] D. Walton. *Argument Evaluation and Evidence*, Springer, 2016.
- [111] D. Walton and T. F. Gordon. Critical question in computational models of legal argument. In Dunne PF and Bench-Capon T (eds). *Argumentation in Artificial Intelligence and Law*,103-111. Wolf Legal Publishers, 2005.
- [112] D. Walton and E.C.W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*, SUNY Series in Logic and Language, State University of New York Press, Albany, 1995.
- [113] J. Williamson and D. Gabbay. Recursive Bayesian Networks and self-fibring logics. In *Laws and models of Science*, D. Gillies, editor, College Publications, 2004, pp 173-247.
- [114] J. Woods, R. H. Johnson, D. M. Gabbay and H. J. Ohlbach. Logic and the Practical turn. In *Handbook of the Logic of Argumentation and Inference: The Turn Towards the Practical*, D. M. Gabbay, R. H. Johnson and H. J. Ohlbach, eds. pp. 1–39. Volume 1 of *Studies in Logic and Practical Reasoning*, North-Holland, 2002.
- [115] J. Woods. Abduction and proof: A criminal paradox,. In Rahman *et al.* (eds), *Approaches to Legal Rationality*, 217-238, Springer, 2010.
- [116] J. Woods. *Errors of Reasoning: Naturalizing the Logic of Inference*, volume 45 of *Studies in Logic*, London: College Publications, 2013. Reprinted with corrections in 2014.
- [117] J. Woods. Inconsistency: Its present impacts and future prospects. In Hewitt and Woods 2015, 158-194.
- [118] J. Woods *Is Legal Reasoning Irrational? An Introduction to the Epistemology of Law*, 2nd ed revised and extended. First edition in 2015. College Publications, 2018.
- [119] J. Woods. *Truth in Fiction: Rethinking its Logic*, Springer, 2018.
- [120] J. Woods. Evidence probativity, and knowledge: A troubled trio. In Hansen HV (ed.) *Proceedings of the Twelfth Meeting of the Ontario Society for the Study of Argumentation*. 2020.

DEFEASIBLE DEONTIC LOGIC: ARGUING ABOUT PERMISSION AND OBLIGATION

HUIMIN DONG

Department of Philosophy (Zhuhai), Sun Yat-sen University, China
huimin.dong@xixilogic.org

BEISHUI LIAO

Department of Philosophy, Zhejiang University, China
baiseliao@zju.edu.cn

RÉKA MARKOVICH, LEENDERT VAN DER TORRE

Department of Computer Science, University of Luxembourg, Luxembourg
{leon.vandertorre,reka.markovich}@uni.lu

Abstract

Defeasible deontic logic uses techniques from non-monotonic logic to address various challenges in normative reasoning, such as prima facie permissions and obligations, moral dilemmas, deontic detachment, contrary-to-duty reasoning and legal interpretation. In this article, we use formal argumentation to design defeasible deontic logics, based on two classical deontic logics. In particular, we use the ASPIC⁺ structured argumentation theory to define non-monotonic variants of well-understood monotonic modal logics. We illustrate the ASPIC⁺-based approach and the resulting defeasible deontic logics using argumentation about strong permission.

1 Introduction to defeasible deontic logic

Deontic logic is the logic of permission, obligation, and prohibition [90], and has been used to formalise reasoning in law, ethics, linguistics, computer science, and elsewhere. See: the deontic logic handbook series [30, 31] for an in depth discussion of this area, the deontic logic textbook [68] for an introduction into the main formal systems, and the handbook of normative multiagent systems [19] for a recent

overview of the challenges in deontic logic and normative reasoning [71], and for an overview of the benchmark examples, inference patterns, and properties [69].

Defeasible deontic logic emerged in the nineteen-nineties when techniques from non-monotonic logic addressed various challenges in normative reasoning. Classical deontic logic is monotonic, meaning that a conclusion derivable from a set of premises remains derivable when new premises are added. However, new premises can block such derivations when normative reasoning involves *prima facie* permissions, conditional permissions, or where one normative principle is preferred to another. Moreover, many axioms of deontic logic have been criticised, and non-monotonic techniques have been applied widely to address them [48, 72, 84]. For instance, Horty [48] formalises normative reasoning in term of conditional obligations. His deontic framework is based on default logic with preference logic. In general, an acceptable derivation may be defeated by a new line of reasoning when new information activates competing normative principles. The Springer volume on defeasible deontic logic [66] appeared over two decades ago, and still provides an excellent overview of challenges in the area of defeasible deontic logic.

Combining *formal argumentation* and deontic logic is an increasingly active research topic in recent years [22, 10, 70]. For example, Prakken [72] proposed combining standard deontic logic with an early-generation formal argumentation system to formalise defeasible deontic reasoning, and Prakken and Sartor [73] formulated arguments about norms as the application of argument schemes to knowledge bases of facts and norms. Young *et al.* [93] proposed an approach to representing prioritised default logic by using the tool ASPIC⁺, and Liao *et al.* [57] represented three logics of prioritised norms using argumentation.

In this article, we use a variant of standard deontic logic [83, 68] as the base logic in an argumentation approach to normative reasoning [26, 81]. The technique proposed here provides a *resolution of conflicts* as a treatment of *prima facie permissions* and *obligations*. For example, conflicts among *prima facie* norms can be resolved using priorities or preferences. The obligation is standard, but we study strong permission instead of weak permission. We argue for choosing strong permission comparing the *moral conflict* pertaining to obligations with that pertaining to permission.

Example 1.1 (Moral Conflict: Obligation). This phenomenon occurs in deontic logic if we reason about deontic dilemmas or conflicts, that is situations where Op and $O\neg p$ both hold. Van der Torre and Tan [88] call this deontic explosion problem “van Fraassen’s paradox”, because van Fraassen [89] gave the following (informal)

analysis of dilemmas in deontic logic. He rejects the AND conjunction pattern,

$$\text{AND: } \frac{O\varphi, O\psi}{O(\varphi \wedge \psi)}.$$

This is because AND warrants a move from the two assumptions $O\varphi$ and $O\neg\varphi$ to the conclusion $O(\varphi \wedge \neg\varphi)$, while such a conclusion is not consistent with the principle ‘ought implies can’ formalised as $\neg O(\varphi \wedge \neg\varphi)$. However, he does not want to reject the conjunction pattern in all cases. In particular, he wants to be able to derive $O(p \wedge q)$ from $O\varphi$ and $O\psi$ when p and q are distinct propositional atoms. His suggestion is that a restriction should be placed on the conjunction pattern: one derives $O(\varphi \wedge \psi)$ from $O\varphi$ and $O\psi$ only if $\varphi \wedge \psi$ is consistent.

Example 1.2 (Moral Conflict: Permission). The sense of moral conflict pertaining to strong permission was first observed by von Wright [90] and later discussed by Lewis [56] and many others [51, 45, 1, 60, 5]. The central property of strong permission can be represented by the following monotonic pattern of free choice permission, FCP, as reviewed recently by Hansson [44]:

$$\text{FCP: } \frac{P\varphi, \Box(\psi \rightarrow \varphi)}{P\psi}$$

The FCP pattern ensures a move from the assumptions $P\phi$ and $\Box(\psi \rightarrow \phi)$ to the conclusion $P\psi$. It then leads from a permission Pp to another permission $P(p \wedge q)$, where q is arbitrary. The moral conflict can arise when q is a proposition bringing moral wrong. An example involving the FCP pattern in natural language [44] is the so-called “vegetarian free lunch” example. In that example, if you are allowed to order a vegetarian lunch, then, by applying the rule `fcP`, you are allowed to order a vegetarian lunch while doing something harmful. In Example 1.1, the moral conflict is brought by obligation aggregation with inconsistency. In contrast, the “paradox” of strong permission here brings up morally wrong statements.

Similar to van Fraassen’s suggestion, we need to restrict the FCP pattern. We accept conclusion $P\psi$ derived from $P\varphi$ and $\Box(\psi \rightarrow \varphi)$ when there is no prohibition on ψ having priority. Compare the following prima facie permissions:

- (A) “It is permitted to use private cars.” (Pc)
- (B) “It is permitted to use private cars which exceed the air pollution level.” ($P(c \wedge a)$)
- (C) “It is permitted to use private cars in an emergency, even those that exceed the air pollution level.” ($P(c \wedge a \wedge e)$)

The FCP pattern guarantees moves from **(A)** to **(B)** and to **(C)**, but the conclusion **(B)** seems to be less acceptable. It is possible to have a prohibition $\neg P(c \wedge a)$ opposing **(B)** and another $\neg P(c \wedge a \wedge e)$ opposing **(C)**. $\neg P(c \wedge a)$ has a higher priority than **(B)** and so **(B)** is defeated. The prohibition $\neg P(c \wedge a \wedge e)$ cannot defeat **(C)**, because this prohibition has a lower priority.

The comparison between Example 1.1 and Example 1.2 suggests that a compromise is required to accept FCP as is required for AND. The counter-intuitive results can be handled with techniques from non-monotonic logic. ASPIC⁺ is one such recent technique that captures the reasoning of normative statements defeasibly. We develop a variety of defeasible deontic logics using ASPIC⁺ in order to model possible reasoning patterns regarding prima facie obligations and permissions.

The layout of this article is as follows. Section 2 provides an overview of various aspects of permission in natural language in legal contexts. Section 3 introduces the running example of this article. Section 4 introduces the basic idea of using formal argumentation as a way to design defeasible deontic logics. Section 5 introduces monotonic deontic logics, and in Section 6 we use these logics to define ASPIC⁺ argumentation systems. Section 7 defines the defeasible deontic logics in terms of the argumentation systems, and Section 8 presents an alternative based on various premises as a further development. In Section 9, we summarise the logical properties of the defeasible deontic logics. Section 10 proposes some further work regarding, for example, related concepts like conditional permissions, rights and duties, permission to know, and permissive norms. We present basic notions regarding permission in various modal languages. By observing the *defeasible* phenomenon in these permissions, we point to possible applications of our ASPIC⁺-based defeasible logics in order to capture their reasoning. Section 11 focuses on related work. Section 12 concludes the article.

2 Many facets of permission

Permission and permissive norms have many facets, proven by the linguistic richness of legal conceptualisation and reasoning in natural language [44, 40]. This section explores various reasoning patterns underlying permission.

One central distinction between different kinds of permissions regards the notions of *declarative* and *descriptive* norms [44], which are two sides of the same coin. A declarative permission is defined by the *presence* of a certain normative, legal, or moral source explicitly granting that permission. By contrast, a descriptive permission can be seen as the *absence* of a *mandatory* source or code containing a prohibition. A declarative permission can generate an obligation, a prohibition, or a

permission. For instance, a declarative permission to the customer, “You are allowed to order your lunch”, generates an obligation on the part of the waiter towards the customer, “I ought to serve the menu”. This kind of permission can sometimes be understood as an explicit, strong, or positive permission [91, 63], because there is a normative source or code that this permission refer to. This is not possible with descriptive permissions. The declarative permission “Every citizen over 18 is allowed to vote” is one instance in the legal context. The civil duty on the state to guarantee the right to vote arises from this permission. Besides, a declarative permission is “action guiding”—the agent would anticipate the deontic status of his or her actions with reference to the permission declared [63]. A phenomenon framed by this effect is the so-called free-choice permission [44]: given that “You are permitted to order a croissant or order a baguette” is declarative, the customer would expect to be allowed to have two choices: the permission to order a croissant and a permission to order a baguette. Otherwise the customer might expect a descriptive permission to order a croissant or might expect a permission to order a baguette, but does not know which one is a permissible option.

Another aspect we need to consider is possible relations between prohibition and permission. We have already stated that a declarative permission is defined by the *presence* of a certain normative, legal, or moral source, and that, by contrast, a descriptive permission can be seen as the *absence* of a *mandatory* source or code. Usually the former is called a strong permission while the latter is called a weak permission [91]. Some may argue that it is not easy to differentiate between strong and weak permissions. In fact, a strong and explicit permission of ϕ can be considered as a free-choice permission—it is action-guiding because of the existence of a norm. Although a strong permission denies a prohibition, we do not consider it to be a weak permission as well. This line follows an idea discussed in several works [91, 1, 44]: it is better not to mix strong and weak permissions. Otherwise, when a permission to ϕ is given, a permission to arbitrary ψ can follow. A strong permission to ϕ can lead to a weak permission to ϕ or ψ , concluding with a strong permission to ψ . Therefore, a clear distinction is required. It is not enough to say that a permission exists because its prohibition is denied. One proper way out is to emphasise which reason supports the existence of such a norm [77, 75], as the next point shows.

We therefore address the third view of permission, which is the main theme of this article, *prima facie* permission. Ross [77] first introduced the notion of *prima facie* with regard to obligations. There may be a moral reason that requires one to do something, which conflicts with another stronger reason for not doing it, and, therefore, the *prima facie* obligation could be defeated. Similarly, a *prima facie* permission could be overtaken by a competing norm when the latter has higher

priority. This competing norm can either be a permission granted to a different person or a prima facie obligation [44]. Here, a prima facie permission is considered to be a declarative and strong permission. It is declarative, because it can further generate an implied permission [44]. A prima facie permission is considered to be a strong permission because it does not necessarily imply that an obligation is denied. A prima facie permission may defeat another permission [44].

Permission is able to represent different types of right in legal theory. The Hohfeldian theory of legal rights [46] usually equates privilege with weak permission [53, 52], while power is created via constitutive rules [78, 79, 64, 29]. We leave further discussion on rights and permissions to Section 10.2.

There are many other linguistic phenomena pertaining to permissions [44, 40], including: unilateral and bilateral permissions; explicit, implied, and tacit permissions; dynamic permissions; permission as activation and revocation; permission as exception of prohibition; and permission as derogation of prohibition. For a more detailed overview, please refer to the *Handbook* chapter on the varieties of permissions [44].

Before turning to our benchmark example of permission, we review the intuition behind, and inferential pattern pertaining to, strong permission in the literature [91, 92, 44]. Von Wright first stated that an action is strongly permitted if “the authority has considered its normative status and decided to permit it” [91, p.68]. Later, he put forward the following pattern that a strong permission should follow: “(Strong) [p]ermission (...) means freedom to choose between all the alternatives, if any, covered by the permitted thing” [92, p.32]. This flavour of “freedom to choose” is similar to the notion of *at liberty* proposed by Raz [75]. It is commonly applied in ordinary language in the following way [45, 5, 7]:

(A) If it is permitted to take a break or continue working, then it is permitted to take a break and it is permitted to continue working.

There are some sentences in the legal context that have a similar sense. We do not take a stand but just present them here: “Exactly how much to tip a server is at the discretion of the customer”, “Bail is granted at the discretion of the court”. It is usually possible to derive some strong permissions to a certain extent from this kind of sentence. To capture sentences like these, as Example 1.2 argues, the FCP inferential pattern of strong permission should be restricted.

3 Running example

In this section, we follow the intuition behind Example 1.2 and present a possible resolution of conflict between prima facie permission and obligation. The principle

of this resolution requires that a prima facie permission must stay consistent with any existing obligation. This intuition can be formalised based on the notion of “Obligation as weak permission”, represented by the OWP rule, which is an axiom in van Benthem’s minimal deontic logic [83]:

$$\text{OWP: } \frac{O\varphi, P\psi}{\Box(\psi \rightarrow \varphi)}$$

It is necessary that what is permitted is not in conflict with what ought to be.

The following example in legal reasoning illustrates the feature of defeasibility displayed by prima facie permission.

1. It is permitted for the owner to use any of his or her property, for example, a private car.
2. It is prohibited to cause air pollution.

The question now is whether it is permitted to use one’s private car which exceeds certain air pollution levels. The solution we adopt is that it is permitted to use private cars in normal situations i.e where the permission is not defeated. So we can derive that:

3. Cars are not used beyond the air pollution level.

If we add information that “this car is used beyond the air pollution level”, or “this car can be used beyond the air pollution level”, and in addition we *prefer* these *specific* statements, we would expect that the derivation of statement (3) will be blocked. Furthermore, when applying the (OWP) rule to assumption (1), the permission to use cars and cause air pollution may be blocked when assumption (2) is preferable. However, by applying the same rule, we would still expect, for example, a permission to use cars while commuting, because there is *no preferable* argument to the contrary.

From a formal point of view, the problem of strong or prima facie permissions that we focus on in this article is the derivation of $P(\varphi \wedge \psi)$ from $P\varphi$. It has been observed by Glavaničová [33] that this is a rule that should not hold in case it leads to inconsistency.

Example 3.1 (Air Pollution). Our aim is to define a defeasible logic such that the prima facie permission to use cars, Pc , can infer a permission to use a car for commuting, $P(c \wedge m)$. Similarly, from Pc we can infer $P(c \wedge a)$, a permission to use cars and cause air pollution. However, in certain exceptional cases, for instance having using cars permitted but having air pollution prohibited, from $\{Pc, O\neg a\}$ we cannot infer $P(c \wedge a)$, and thus the logic is non-monotonic.

To capture the non-monotonic reasoning in Example 3.1, we adopt a variant of van Benthem’s [83] minimal deontic logic as the monotonic base. This logic contains two axioms for the FCP and OWP patterns, which are necessary for the derivations in Example 3.1.

Each level in our approach can be analysed using the methods pertaining to the relevant discipline, i.e. monotonic logic can be studied using, for example, modal logics based on possible world semantics); argumentation theory can be studied using rationality postulates [16]; and non-monotonic inference can be analysed using, for example, the approach advocated by Kraus et al. [54].

4 Using ASPIC+ to design defeasible deontic logics

In this article, we explain the basic idea of employing ASPIC+ to design deontic argumentation systems and defeasible deontic logics and, in particular, to study strong permission. The ASPIC+ approach has been discussed in a variety of papers [81, 26]. In our opinion, this approach is one of the most transparent ones suitable to put forward our idea in argumentation, and we follow the exposition provided by Modgil and Prakken in the handbook of formal argumentation [65]. ASPIC+ is a framework for specifying argumentation systems, and it leaves one full freedom to choose the logical language, the strict and defeasible inference rules, the axioms and ordinary premises in a knowledge base, and the argument preference ordering [65].

Modgil and Prakken [65] observe that “in ASPIC+ and its predecessors, going back to the seminal work of John Pollock, arguments can be formed by combining strict and defeasible inference rules and conflicts between arguments can be resolved in terms of a preference relation on arguments. This results in abstract argumentation frameworks (a set of arguments with a binary relation of defeat), so that arguments can be evaluated with the theory of abstract argumentation.” In this article, we use argumentation systems to define defeasible deontic logics. Our ASPIC+-based methodology consists of three steps.

1) Arguments: we take literally Modgil and Prakken’s [65] idea that “rule-based approaches in general do not adopt a single base logic but two base logics, one for the strict and one for the defeasible rules”. We use monotonic modal logics as our base logics with Hilbert-style proof theory.

1.1) Strict arguments use only strict rules defined in terms of a “lower bound” logic that defines the minimal inferences that must be made. We use a variant of von Wright’s standard deontic logic [68] for strict arguments.

- 1.2) **Defeasible arguments** also use defeasible rules defined in terms of an “upper bound” logic that defines all possible inferences that can be made. We use a variant of van Benthem’s logic of strong permission [83] for defeasible arguments.
- 2) **Preferences among arguments** can be generic or can depend on the logical languages used to build the arguments. We focus on **argument types** defined in ASPIC⁺ that distinguish between defeasible and plausible arguments.
- 3) **Non-monotonic inference relations** can be based on a sceptical or credulous relation, and on one of the argumentation semantics. Here, we only choose the sceptical inferential relation based on stable semantics.

In the following sections, we present the above notions in ASPIC⁺ step by step.

5 Arguments based on two monotonic logics

We use two monotonic logics to define the strict and defeasible rules in ASPIC⁺, and use the crude approach to define arguments [65]: “A crude way is to simply put all valid propositional (or first-order) inferences over your language of choice in [the strict rules] R_s . So if a propositional language has been chosen, then R_s can be defined as follows (where \vdash_{PL} denotes standard propositional-logic consequence). For any finite $S \subseteq \mathcal{L}$ and any $\phi \in \mathcal{L}$: $S \rightarrow \phi \in R_s$ if and only if $S \vdash_{PL} \phi$.” This method can be applied to define defeasible rules, and this application, as stated by Modgil and Prakken [65], should be based on some cognitive or rational criteria. By using the crude method to define strict rules in the lower-bounded logic \mathbf{S}^- and to define defeasible rules in the upper-bounded logic \mathbf{S}^+ , the arguments can be short even when Hilbert style derivations are quite long.

Besides this way of defining the defeasible rules, all the other definitions in this section—like the arguments and the extensions—are standard and taken from the *Handbook* chapter by Modgil and Prakken [65]. In particular, we consider three instantiations of ASPIC⁺ by taking different monotonic logics (\mathbf{D}_{-1} or \mathbf{D}_{-2} , defined later) as the basic logic and then treating either as only FCP or FCP together with OWP (in Table 1) as defeasible. In this section, we define the notion of argumentation theory. In the following section, we use argumentation theory to define non-monotonic logic as a combination of two selected monotonic logics: \mathbf{S}^- , \mathbf{S}^+ .

We first present a version of van Benthem’s [83] deontic logic of obligation and strong permission. This logic is different from standard deontic logic [68]. Standard deontic logic sees obligation and permission as a dual pair representing that what is permitted is not obligatory not to be. Van Benthem’s deontic logic does not take this

view. While this logic still interprets obligation as what is necessary for staying ideal, it interprets permission as what is sufficient for staying ideal. This new connection can be formalised as OWP. The modal language contains the classic negation \neg , conjunction \wedge , universal modality \Box , and two additional deontic modalities: O for obligation and P for strong permission.

Definition 5.1 (Deontic Language). Let p be any element of a given (countable) set $Prop$ of atomic propositions. The deontic language \mathcal{L} of modal formulas is defined as follows:

$$\phi := p \mid \neg\phi \mid (\phi \wedge \psi) \mid \Box\phi \mid O\phi \mid P\phi$$

The disjunction \vee , the material implication \rightarrow and the existential modality \Diamond are defined as usual: $\phi \vee \psi := \neg(\neg\phi \wedge \neg\psi)$, $\phi \rightarrow \psi := \neg(\phi \wedge \neg\psi)$ and $\Diamond\phi := \neg\Box\neg\phi$.

Definition 5.2. The deontic logic \mathbf{D} is a system that includes all the axioms and rules in Table 1.

The axiomatisation presented in Table 1 is a variant of van Benthem’s logic [83]. We use \mathbf{D} to denote it. The deontic logic \mathbf{D} not only takes obligation and universal modality into account, but also considers free-choice permission and the connection between obligation and permission. In logic \mathbf{D} , except for the essential K_\Box , E_\Box , T_\Box , 4_\Box , B_\Box , and NEC_\Box (NEC stands for necessity), the axioms \Box_O and \Box_P are the core of the universal modality in normal modal logic. Moreover, \Box_O claims that what is always the case is obligatory, but \Box_P leaves the space for what is never to be permitted. The axiom D_O maintains that an obligation is to be ideally consistent as usual. OWP considers “obligation as the weakest permission” [83, 3]. RFC (it stands for “Reverse of free choice permission”) represents one direction of free-choice permission, and FCP the other. For further information about the logic and its motivations, see the work of van Benthem [83].

In this article, we consider sub-systems of \mathbf{D} that contain a strict subset of the axioms and inference rules of \mathbf{D} . In particular, we define \mathbf{D}_{-1} as an axiomatisation that does not contain FCP, and we define \mathbf{D}_{-2} as an axiomatisation that does not contain OWP and other axioms (FCP, RFC and \Box_P) used purely for permission.

Definition 5.3. The deontic logic \mathbf{D}_{-1} is a system that includes all the axioms and rules in \mathbf{D} except FCP. The deontic logic \mathbf{D}_{-2} is a system that includes all the axioms and rules in \mathbf{D} except RFC, \Box_P , FCP, and OWP.

We define the notions of derivation based on modal logic $\mathbf{S} \in \{\mathbf{D}, \mathbf{D}_{-1}, \mathbf{D}_{-2}\}$ in the usual way, see [14] for instance. Note that modal logic provides two related

- PL: all propositional tautologies	- K_{Δ} : $\Delta(\phi \rightarrow \psi) \rightarrow (\Delta\phi \rightarrow \Delta\psi)$
- E_{\square} : $\square\phi \leftrightarrow \neg\diamond\neg\phi$	- T_{\square} : $\square\phi \rightarrow \phi$
- 4_{\square} : $\square\phi \rightarrow \square\square\phi$	- B_{\square} : $\phi \rightarrow \square\diamond\phi$
- \square_O : $\square\phi \rightarrow O\phi$	- \square_P : $P\perp$
- D_O : $\neg(O\phi \wedge O\neg\phi)$	- OWP: $O\phi \wedge P\psi \rightarrow \square(\psi \rightarrow \phi)$
- RFC: $P\phi \wedge P\psi \rightarrow P(\phi \vee \psi)$	- FCP: $P\psi \wedge \square(\phi \rightarrow \psi) \rightarrow P\phi$
- MP: $\phi, \phi \rightarrow \psi / \psi$	- NEC_{Δ} : $\phi / \Delta\phi$
where $\Delta \in \{\square, O\}$	

 Table 1: The logic **D** of obligation and permission

kinds of derivation according to the application of necessitation, i.e. necessitation can only be applied to theorems and not to an arbitrary set of formulas. We use both notions in the formal argumentation theory.

Definition 5.4 (Derivations without Premises). Let $\mathbf{S} \in \{\mathbf{D}, \mathbf{D}_{-1}, \mathbf{D}_{-2}\}$ be a deontic logic. A derivation of ϕ in \mathbf{S} is a finite sequence $\phi_1, \dots, \phi_{n-1}, \phi_n$ such that $\phi = \phi_n$, and for every $\phi_i (1 \leq i \leq n)$ in this sequence, ϕ_i is

1. either an instance of one of the axioms in \mathbf{S} , or
2. the result of the application of one of the rules in \mathbf{S} to those formulas appearing before ϕ_i .

We write $\vdash_{\mathbf{S}} \phi$ if there is a derivation of ϕ in \mathbf{S} , or, $\vdash \phi$ when the context of \mathbf{S} is clear. We say ϕ is a theorem of \mathbf{S} , or \mathbf{S} proves ϕ . We write $Cn(\mathbf{S})$ to represent the set of all the theorems of \mathbf{S} .

Definition 5.5 (Derivations from Premises). Let $\mathbf{S} \in \{\mathbf{D}, \mathbf{D}_{-1}, \mathbf{D}_{-2}\}$ be a deontic logic. Given a set Γ of formulas, a derivation of ϕ from Γ in \mathbf{S} is a finite sequence $\phi_1, \dots, \phi_{n-1}, \phi_n$ such that $\phi = \phi_n$, and for every $\phi_i (1 \leq i \leq n)$ in this sequence, ϕ_i is

1. either $\phi_i \in Cn(\mathbf{S}) \cup \Gamma$; or
2. the result of the application of one of the rules (which is neither NEC_{\square} nor NEC_O) to those formulas appearing before ϕ_i .

We write $\Gamma \vdash_{\mathbf{S}} \phi$ if there is a derivation from Γ for ϕ in \mathbf{S} ¹, or, $\Gamma \vdash \phi$ when the context of \mathbf{S} is clear. We say that ϕ is derivable in \mathbf{S} from Γ . We write $Cn_{\mathbf{S}}(\Gamma)$ to

¹Alternatively, it can be seen as a theorem $\vdash_{\mathbf{S}} \bigwedge \Gamma \rightarrow \phi$ by the so-called deduction theorem.

represent the set of formulas derivable in \mathbf{S} from Γ , or $Cn(\Gamma)$ if the context of \mathbf{S} is clear.

A system \mathbf{S} is consistent iff $\perp \notin Cn(\mathbf{S})$; otherwise, it is inconsistent. A set Γ is \mathbf{S} -consistent iff $\perp \notin Cn_{\mathbf{S}}(\Gamma)$; otherwise, it is inconsistent. A set $\Gamma' \subseteq \Gamma$ is a maximally \mathbf{S} -consistent subset of Γ , denoted as $\Gamma' \in MC_{\mathbf{S}}(\Gamma)$, iff there is no $\Gamma'' \supset \Gamma'$ such that Γ'' is \mathbf{S} -consistent.

The following example explains in what sense we can say in monotonic logics that Pc and $O\neg a$ are in conflict. These two assumptions will not be consistent when taken together with the statement “It is not the case that using a car does not lead to air pollution”, i.e. $\neg\Box(c \rightarrow \neg a)$, which can be equally formalised as $\Diamond(c \wedge a)$. This will be explained in Example 5.1.

Example 5.1 (Air Pollution, continued). The following derivation shows that the set $\{Pc, O\neg a, \Diamond(c \wedge a)\}$ is inconsistent in \mathbf{D}_{-1} or \mathbf{D} .

- | | |
|--|-----------------|
| 1. $O\neg a \wedge \Diamond(c \wedge a)$ | assumptions, PL |
| 2. $O\neg a \wedge Pc \rightarrow \Box(c \rightarrow \neg a)$ | OWP |
| 3. $\Diamond(c \wedge a) \leftrightarrow \neg\Box(c \rightarrow \neg a)$ | E_{\Box} |
| 4. $O\neg a \wedge \Diamond(c \wedge a) \rightarrow \neg Pc$ | 2, 3, MP |
| 5. $\neg Pc$ | 1, 4, MP |

To investigate when *defeasible* derivations are possible, we use \mathbf{D} only to derive conclusions that are defeasible, and we use one of the subsystems of \mathbf{D} to define monotonic and strict conclusions.

We assume the view of ASPIC⁺ of considering inference rules as uncertain and fallible defeasible rules, and those rules that are infallible as strict rules. This type of uncertainty or fallibility is represented by distinguishing between lower-bounded and upper-bounded logics. However, to simplify the present issue of how to use ASPIC⁺ to define non-monotonic logics, it is not necessary to fully adopt all the methods in ASPIC⁺ to define arguments. We only consider a general knowledge base here. The distinction between different types of knowledge is left until Section 8.

Definition 5.6 (Argumentation Theory). Let \mathcal{L} be the deontic language and let $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a Cartesian product of two monotonic logics. An argumentation system AS based on $(\mathbf{S}^-; \mathbf{S}^+)$ is a pair (\mathcal{L}, R) where $R = R_s \cup R_d$ is a set of rules such that:

- $R_s = \{\phi_1, \dots, \phi_n \mapsto \phi \mid \{\phi_1, \dots, \phi_n\} \vdash_{\mathbf{S}^-} \phi\}$ is the set of strict rules, and
- $R_d = \{\phi_1, \dots, \phi_n \Rightarrow \phi \mid \{\phi_1, \dots, \phi_n\} \vdash_{\mathbf{S}^+} \phi \ \& \ \{\phi_1, \dots, \phi_n\} \not\vdash_{\mathbf{S}^-} \phi\}$ is the set of defeasible rules.

If the context of $(\mathbf{S}^-; \mathbf{S}^+)$ is clear, we mention AS without $(\mathbf{S}^-; \mathbf{S}^+)$. An argumentation theory AT is a pair (AS, K) where $K \subseteq \mathcal{L}$ is a knowledge base.

So the requirement that $R_s \cap R_d = \emptyset$ holds. We define the sets of empty-bodied strict/defeasible rules as $R_s^0 = \{\mapsto \phi \mid \mapsto \phi \in R_s\}$ and $R_d^0 = \{\rightrightarrows \phi \mid \rightrightarrows \phi \in R_d\}$. Clearly, $R_s^0 \subseteq R_s$ and $R_d^0 \subseteq R_d$.

Next, we define what are arguments. We will see that arguments have different structures to those of derivations. Although each argument corresponds to a derivation defined as a top rule, the former explicitly considers each step of this derivation as a finite sequence.

Definition 5.7 (Arguments). Let AT be an argumentation theory with a knowledge base K and an argumentation system (\mathcal{L}, R) . Given each $n \in \mathbb{N}$, the set \mathcal{A}_n where $n \in \mathbb{N}$ is defined as follows:

$$\begin{aligned} \mathcal{A}_0 &= K \cup R_s^0 \cup R_d^0 \\ \mathcal{A}_{n+1} &= \mathcal{A}_n \cup \{B_1, \dots, B_m \triangleright \psi \mid B_i \in \mathcal{A}_n \text{ for all } i \in \{1, \dots, m\}\} \end{aligned}$$

where $\triangleright \in \{\mapsto, \rightrightarrows\}$, and for an element $B \in \mathcal{A}_i$ with $i \in \mathbb{N}$:

- if $B = \psi \in K$, then $Prem(B) = \{\psi\}$, $Conc(B) = \psi$, $Sub(B) = \{\psi\}$, $Rules_d(B) = \emptyset$, and $TopRule(B) = \text{undefined}$;
- if $B = \mapsto \psi \in R_s^0$, then $Prem(B) = \emptyset$, $Conc(B) = \psi$, $Sub(B) = \{\mapsto \psi\}$, $Rules_d(B) = \emptyset$, and $TopRule(B) = \mapsto \psi$;
- if $B = \rightrightarrows \psi \in R_d^0$, then $Prem(B) = \emptyset$, $Conc(B) = \psi$, $Sub(B) = \{\rightrightarrows \psi\}$, $Rules_d(B) = \{\rightrightarrows \psi\}$, and $TopRule(B) = \rightrightarrows \psi$;
- if $B = B_1, \dots, B_m \triangleright \psi$ where \triangleright is \mapsto , then $\{Conc(B_1), \dots, Conc(B_m)\} \mapsto \psi \in R_s$ with $Prem(B) = Prem(B_1) \cup \dots \cup Prem(B_m)$, $Conc(B) = \psi$, $Sub(B) = Sub(B_1) \cup \dots \cup Sub(B_m) \cup \{B\}$, $Rules_d(B) = Rules_d(B_1) \cup \dots \cup Rules_d(B_m)$, $TopRule(B) = Conc(B_1), \dots, Conc(B_m) \mapsto \psi$; and
- if $B = B_1, \dots, B_m \triangleright \psi$ where \triangleright is \rightrightarrows , then each condition is similar to the previous item, except that the rule is defeasible and $Rules_d(B) = Rules_d(B_1) \cup \dots \cup Rules_d(B_m) \cup \{Conc(B_1), \dots, Conc(B_m) \rightrightarrows \psi\}$.

We define $\mathcal{A} = \bigcup_{n \in \mathbb{N}} \mathcal{A}_n$ as the set of arguments on the basis of AT , and define $Conc(E) = \{\varphi \subseteq Conc(A) \mid A \in E\}$ where $E \subseteq \mathcal{A}$. Let $F(B) = Conc(Sub(B))$ when $B \in \mathcal{A}$. We have $F(E) = \bigcup \{F(B) \mid B \in E \subseteq \mathcal{A}\}$.

The following example illustrates the arguments provided in the running example. We consider the defeats (arrows) in Figure 2 in the next section.

Example 5.2 (Air Pollution, continued). Let $K = \{\diamond(c \wedge a), O\neg a, Pc\}$ be a knowledge base where the atomic proposition c stands for someone using cars, and atomic proposition a stands for someone causing air pollution. Prohibition or forbidden means “ought not to”. There are three arguments in knowledge base K :

- $A = O\neg a$: It is prohibited for someone to cause air pollution;
- $B = Pc$: It is permitted for someone to use cars;
- $C = \diamond(c \wedge a)$: It is possible for someone who uses cars to cause air pollution.

Some arguments constructed from K are shown as follows:

1. the arguments that have top rules as strict rules by using NEC_O and K_O in \mathbf{D}_{-2} :

- $A''' = A \mapsto O\neg(c \wedge a)$

2. the arguments that have top rules as defeasible rules by using OWP and E_{\square} in \mathbf{D}_{-1} :

- $A' = A, C \Rightarrow \neg Pc$
- $A'' = A, C \Rightarrow \neg P(c \wedge a)$
- $B' = B, C \Rightarrow \neg O\neg a$
- $B''' = B, C \Rightarrow \neg O\neg(c \wedge a)$
- $C' = A, B \Rightarrow \neg \diamond(c \wedge a)$

3. the arguments that have top rules as defeasible rules by using FCP in \mathbf{D}_{-1} :

- $B'' = B \Rightarrow P(c \wedge a)$
- $B''' = B \Rightarrow P(c \wedge m)$

where m is short for someone commuting.

Because arguments A' and A''' both have premise $O\neg a$, we consider that these two arguments represent this obligation $O\neg a$. On the other hand, arguments B' and B'' both have Pc as a premise, and we consider that they represent permission Pc .

The formulas pertaining to the set $\{A'', A''', B'''\}$ of arguments are $\{O\neg a, O\neg(c \wedge a), \diamond(c \wedge a), Pc, \neg P(c \wedge a), P(c \wedge m)\}$. In Section 6, we present a mechanism for selecting this desired set of arguments, so that the defeasible deontic logic corresponds to it.

6 Preferences among arguments

In this article, we follow the idea proposed by Modgil and Prakken [65] of partitioning arguments on the basis of strict, defeasible, and sound arguments. These partitions on arguments can be used to define two orders: rule-based and premise-based. The rule-based order prefers strict arguments to defeasible arguments, while premise-based order prefers unsound arguments to sound ones.

Definition 6.1 (Argument Properties). Let A, B be arguments and E a set of arguments. Then A is strict if $Rules_d(A) = \emptyset$, it is defeasible if $Rules_d(A) \neq \emptyset$, and it is sound if $Prem(A) \cap K \neq \emptyset$. We define $Concs(E) = \{Conc(A) \mid A \in E\}$. The partial order \leq over E is rule-based iff we have $A \leq B$ iff A is defeasible, and it is premise-based iff $A \leq B$ iff A is sound.

We use \leq^τ to denote a τ -ordering with $\tau \in \{r, p\}$, where r stands for rule-based and p stands for premise-based. The premise-based ordering \leq is an *universal* order, because given any $A, B \in \mathcal{A}$, it is the case that $A \leq B$.

Next, we introduce the notions of defeat. The first notion is rebuttal and the second is undermining [65]. In order to simplify the discussion, we do not make any additional assumptions like distinguishing between different kinds of defeated knowledge on undermining. In the next section, distinguishing between rebuttal and undermining will give different consequences in defeasible deontic reasoning.

Definition 6.2 (Argumentation Frameworks). Given $A, B \in \mathcal{A}$ and an order \leq over \mathcal{A} , argument A defeats argument B , or simply call it a *defeat*, denoted as $(A, B) \in \mathcal{D}$ if and only if:

- A *rebutts* B : $Conc(A) = +\phi$ for some $B' \in Sub(B)$ and $TopRule(B') \in R_d$, $Conc(B') = -\phi$, and $A \not\leq B'$, or
- A *undermines* B : $Conc(A) = +\phi$ for knowledge $-\phi \in Prem(B)$ of B and $A \not\leq -\phi$,

where $+\phi$ indicates m negations in front of ϕ , and $-\phi$ indicates n negations in front of ϕ , such that $|m - n|$ is an odd number. An abstract argumentation framework AF corresponding to $\langle AT, \leq^\tau \rangle$ is a pair $(\mathcal{A}, \mathcal{D})$ where \mathcal{D} is the set of all defeats defined by \leq over \mathcal{A} .

As the following example shows, the notions of defeat can explain the idea of one rule taking precedence over another. Notice that knowledge, defeasible rules, and preferences are the three key elements to deciding what are defeated.

Example 6.1 (Defeats in knifed murder, continued). As shown in Figure 1, when in the rule-based ordering, A''' rebuts B''' , but not vice versa. This shows a case of obligation defeating permission but not vice versa.

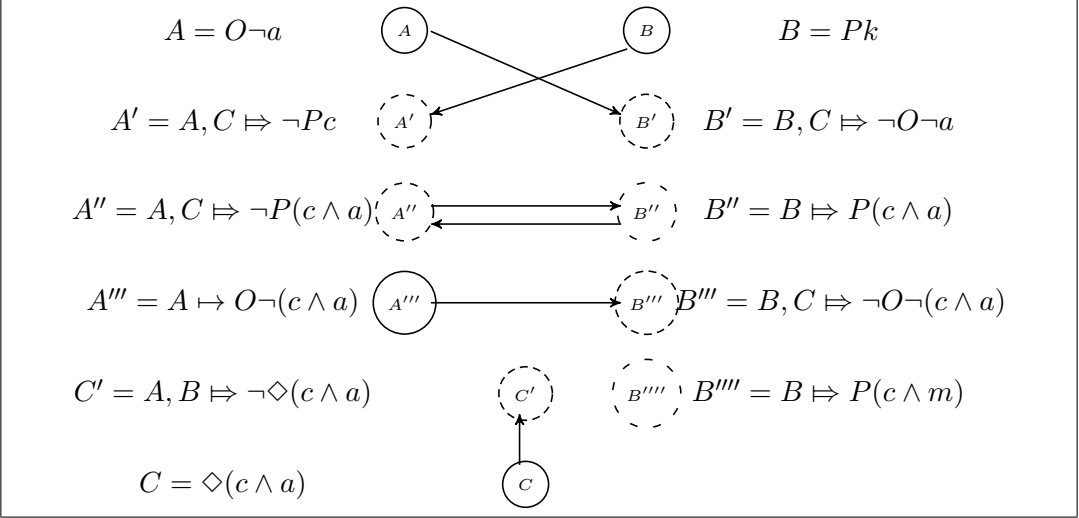


Figure 1: Some of the defeats among arguments based on $(\mathbf{D}_{-2}; \mathbf{D})$ in the rule-based ordering. Straight arrows are defeats among these arguments.

Now we turn to the premise-based ordering. As shown in Figure 2, we have that B' undermines A''' , which indicates a permission defeating an obligation. Here, we can see that all rebuttals in the rule-based ordering are also maintained in the premise-based ordering. So the straight arrows in Figure 2 represent defeat relations under the rule-based ordering, and the dashed arrows represent the additional defeat relations under the premise-based ordering.

Definition 6.3 (Dung Extensions). Let $AF = (\mathcal{A}, \mathcal{D})$ and let $E \subseteq \mathcal{A}$ be a set of arguments. Then:

- E is conflict-free iff $\forall A, B \in E$, we have $(A, B) \notin \mathcal{D}$;
- $A \in \mathcal{A}$ is acceptable w.r.t. E iff when $B \in \mathcal{A}$ such that $(B, A) \in \mathcal{D}$, then $\exists C \in E$ such that $(C, B) \in \mathcal{D}$;
- E is an admissible set iff E is conflict-free, and if $A \in E$, then A is acceptable w.r.t. E ;
- E is a complete extension iff E is admissible, and if $A \in \mathcal{A}$ is acceptable w.r.t. E , then $A \in E$;

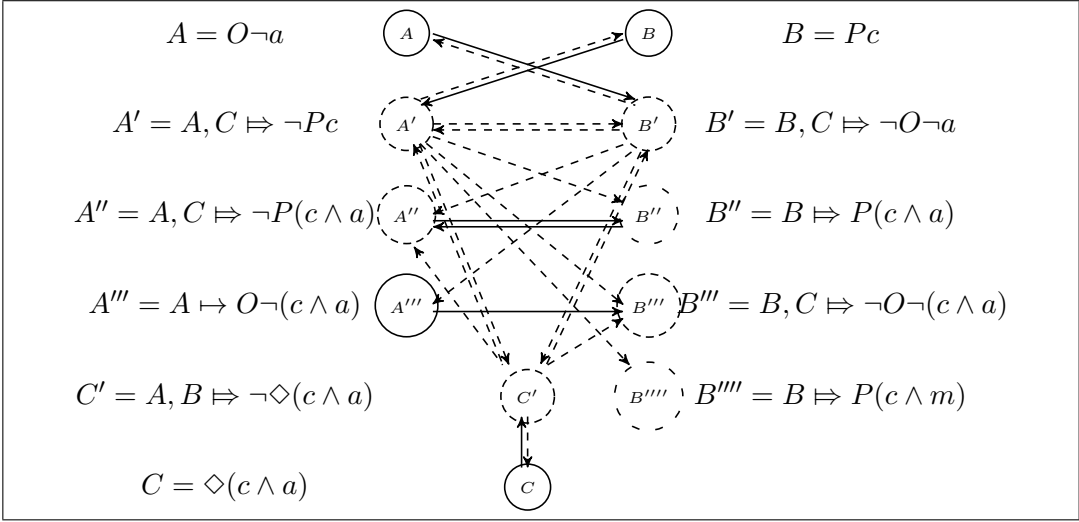


Figure 2: Some of the defeats among arguments based on $(\mathbf{D}_{-2}; \mathbf{D})$ in the premise-based ordering. Straight arrows are rebuttals while dashed arrows are underminings.

- E is a stable extension iff E is conflict-free and $\forall B \notin E \exists A \in E$ such that $(A, B) \in \mathcal{D}$.

The following example illustrates a different sense of consistency in ASPIC⁺ by using stable extensions in order to explain, given the inconsistent knowledge base K , why $B \Rightarrow P(c \wedge a)$ is sometimes defeated and why $B \Rightarrow P(c \wedge m)$ is always undefeated.

Example 6.2 (Air Pollution, continued). Consider the arrows in Figure 1. The straight arrows represent defeat relations under the rule-based ordering, and the dashed arrows represent additional defeat relations under the premise-based or universal ordering. Under the rule-based ordering, arguments A , B and C will not be defeated in every extension, whereas in premise-based or universal ordering, they will be. For this reason, we prefer the rule-based ordering in this example. Furthermore, under the rule-based ordering, we have at least two stable extensions, one containing $B \Rightarrow P(c \wedge a)$ and another containing $A, C \Rightarrow \neg P(c \wedge a)$. Since $B'''' = B \Rightarrow P(c \wedge m)$ will not be defeated, we have B'''' in every stable extension. Similarly, arguments in the form of $A_1, \dots, A_n \mapsto Pk \vee O\neg a \vee \diamond(c \wedge a)$ are contained in every stable extension.

Not only can plausible and defeasible arguments be compared in the preference ordering, factual statements can be preferred to deontic statements, and prohibitions to permissions or vice versa. We leave such further investigations until Section 8.

7 Designing defeasible deontic logics

Our defeasible deontic logics are designed by using the stable extensions pertaining to different monotonic logics and different orderings. The proposition that follows provides a guideline for searching for these stable extensions. In the case of premise-based ordering, strict rules are equally preferable to defeasible rules. So a stable extension can be considered as a maximally consistent subset of knowledge base K in the upper-bounded logic. We call this an *undermining*-based construction, for details see e.g. [4, 81]. But this is not enough to capture the case of rule-based ordering in which the defeasible argument is less preferable compared to the others. So the second item of this proposition provides a rule-based method for constructing the desired extensions, stable extensions. We construct each stable extension in the style of Lindenbaum’s Lemma [14]. That is, we first consider the maximally consistent subset K' of the knowledge base with regard to the lower-bounded logic \mathbf{S}^- for strict rules, and then a consistent subset of K' with regard to the upper-bounded logic \mathbf{S}^+ for defeasible rules, such that no argument with regard to \mathbf{S}^+ defeats that with regard to \mathbf{S}^- , and K' is a maximal set satisfying these two conditions. This is called a *rebuttal*-based construction. It can be considered as a way of fibring—combining two logics [32]. See the following proposition for details.

Proposition 7.1. Consider the deontic language \mathcal{L} and a pair of two monotonic logics $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$. Let AF , corresponding to $\langle AT, \leq^\tau \rangle$, be an abstract argumentation framework $(\mathcal{A}, \mathcal{D})$ such that AT is based on $(\mathbf{S}^-; \mathbf{S}^+)$, K is a knowledge base, and $\tau \in \{p, r\}$. Given a set $\Gamma \subseteq \mathcal{L}$ of formulas, we define:

- a stable set generated by Γ as $\{D \in \mathcal{A} \mid F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma)\}$;
- a proper set generated by Γ as $\bigcup_{i \in \omega} E_i$, such that

$$E_0 = \{D \in \mathcal{A} \mid F(D) \subseteq Cn_{\mathbf{S}^-}(\Gamma)\}$$

$$E_{n+1} = \begin{cases} E_n \cup \{D \in \mathcal{A}\}, & \text{if } F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma) \text{ and} \\ & F(D) \cup F(E_n) \text{ is } \mathbf{S}^- \text{-consistent;} \\ E_n, & \text{otherwise.} \end{cases}$$

1. When $\tau = p$, then E is a stable set generated by $\Gamma \in MC_{\mathbf{S}^+}(K)$ iff E is a stable extension regarding K .
2. When $\tau = r$, E is a proper set generated by $\Gamma \in MC_{\mathbf{S}^-}(K)$ iff E is a stable extension regarding K .

Given the knowledge base $K = \{Pc, O\neg a, \diamond(c \wedge a)\}$ of the running example, ASPIC⁺ provides a mechanism for deciding whether the two arguments $A''' = A \mapsto O\neg(c \wedge a)$ and $B''' = B, C \Rightarrow \neg O\neg(c \wedge a)$ can be accepted. In the case of premise-based order, undermining together with stability is a mechanism for ensuring that even when knowledge base K is not consistent, there is still a way to find maximally consistent subsets to construct stable extensions. In the case of rule-based ordering, we cannot use the undermining-based construction to ensure that we derive the first argument but not the second one. Instead, we need to use the rebuttal-based construction. The rebuttal approach can accept both the above arguments, unless one works contrary to the other. That is why the two arguments need to be distinguished in the lower-bounded and upper-bounded logics.

We now present the central definition of the article, namely the definition of defeasible deontic logic in terms of formal argumentation theory. This is well in line with current practice in ASPIC⁺. We first take the desired conclusions in each stable extension (as shown in Proposition 7.1) and then the intersection of all the stable extensions.

Definition 7.1 (Defeasible Inferences). Let $\Gamma \subseteq \mathcal{L}$ and $\phi \in \mathcal{L}$. We let $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a Cartesian product of two monotonic logics, and let \leq^τ be a τ -ordering such that $\tau \in \{r, p\}$. Let AT be a Γ -argumentation theory based on $(\mathbf{S}^-; \mathbf{S}^+)$ iff the argumentation theory AT obtains with $K = \Gamma$, and iff $AF^\tau = \langle AT, \leq^\tau \rangle$. The non-monotonic inference $|\sim_{\mathbf{S}^-; \mathbf{S}^+}^\tau$ is defined as follows:

- $\Gamma |\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \vee} \phi$ iff every stable extension of the Γ - AT based on $(\mathbf{S}^-; \mathbf{S}^+)$ corresponding to AF^τ contains an argument A with $Conc(A) = \phi$.

We define the closure operator corresponding to this inference relation as usual: $\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \vee}(\Gamma) = \{\phi \mid \Gamma |\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \vee} \phi\}$. Moreover, we write $|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \vee} \phi$ when $\emptyset |\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \vee} \phi$.

The resulting non-monotonic inference relations are standard relations among sets of formulas pertaining to the logical language, i.e. they no longer refer to ASPIC⁺. An alternative way to define non-monotonic logics is to first consider the intersection of all stable extensions, and then the conclusions of the arguments that appear in the intersection. For instance, $Pc \vee O\neg a \vee \diamond(c \wedge a)$ is an element in $\mathcal{C}_{\mathbf{D}_{-2}; \mathbf{D}}^{\tau \vee}(\{Pc, O\neg a, \diamond(c \wedge a)\})$ where $\tau \in \{p, r\}$. With the alternative approach mentioned above, this cannot be inferred because it is possible to have many different arguments, for instance $Pc \Rightarrow Pc \vee O\neg a \vee \diamond(c \wedge a)$ and $O\neg a \Rightarrow Pc \vee O\neg a \vee \diamond(c \wedge a)$, that contain the same conclusion but from different premises.

The following proposition offers a detailed explanation of the mechanisms we have proposed. First, the undermining mechanism states that the non-monotonic

consequences are the intersection of all maximally consistent subsets of the knowledge base under an universal or premise-based ordering. Second, and more generally, the rebuttal mechanism states that the non-monotonic consequences are encased in all unions of a maximally consistent subset of the knowledge base with regard to the lower-bounded logic, and are encased in a consistent subset of those unions with regard to the upper-bounded logic in certain maximal behaviour.

Proposition 7.2. Let $\Gamma \subseteq \mathcal{L}$, $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics and let K be a knowledge base of AT . We define

- an R-set generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$ as $\bigcup_{n \in \mathbb{N}} R_n$, such that

$$R_0 = Cn_{\mathbf{S}^-}(\Gamma)$$

$$R_{n+1} = \begin{cases} R_n \cup \{\varphi\}, & \text{if } \varphi \in Cn_{\mathbf{S}^+}(\Gamma) \text{ and} \\ & \{\varphi\} \cup R_n \text{ is } \mathbf{S}^- \text{-consistent;} \\ R_n, & \text{otherwise;} \end{cases}$$

where $\Gamma \in MC_{\mathbf{S}^-}(K)$.

The R-collection $R_{\mathbf{S}^-; \mathbf{S}^+}(K)$ generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$ is the set of all R-sets generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$. Then:

1. $C_{\mathbf{S}^-; \mathbf{S}^+}^{p\forall}(K) = \bigcap_{\Gamma \in MC_{\mathbf{S}^+}(K)} Cn_{\mathbf{S}^+}(\Gamma)$;
2. $C_{\mathbf{S}^-; \mathbf{S}^+}^{r\forall}(K) = \bigcap R_{\mathbf{S}^-; \mathbf{S}^+}(K)$.

To prove Proposition 7.2, and inspired by Proposition 7.1, we first consider the maximally consistent subset of the knowledge base with regard to the lower-bounded logic \mathbf{S}^- , and then consider the consistent subset of the knowledge base with regard to the upper-bounded logic \mathbf{S}^+ , such that the second consistent set is maximal in the sense that it is consistent with each element of the first consistent set with regard to the lower-bounded logic. In contrast, Proposition 7.2.2 illustrates a different understanding of maximality of consistency, which not only has to consider the consistency of the upper-bounded logic but also its consistency with each element in the lower-bounded logic. Our method of defining defeasible deontic logic follows from the “layer” method, which has been used to deal with “paraconsistency” [13]. These methods share a similar spirit of handling maximal consistency when instantiating formal argumentation based on classical logic [2] or modal logic [10]. What we have done here is to explicitly construct a stable extension according to the variants on the upper-/lower-bounded logics as well as the variants on the orders.

We leave to further research a formal analysis of the non-monotonic inference relation, as well as the development of alternative non-monotonic relations in terms of the formal argumentation theory.

Example 7.1 (Air Pollution, continued). Given a knowledge base $K = \{\diamond(c \wedge a), O\neg a, Pc\}$ as the premises, we have different non-monotonic consequences shown in Table 2, depending on the combinations of monotonic logics and orderings. They are non-monotonic in the sense that, even given Pk as one premise, $P(c \wedge a)$ is excluded in every non-monotonic consequence, while $P(c \wedge m)$ is a non-monotonic consequence with regard to $(\mathbf{D}_{-2}; \mathbf{D})$ under the rule-based ordering. Intuitively speaking, in Figure 1 there is no defeat of arguments ending with $P(c \wedge m)$, while there is an argument A'' that defeats a B'' that ends with $P(c \wedge a)$.

	Order	Closure	Example of Consequences
$(\mathbf{D}_{-2}; \mathbf{D}_{-1})$	p	T_p	$\bigvee K$
$(\mathbf{D}_{-2}; \mathbf{D}_{-1})$	r	T_r^1	$\diamond(c \wedge a), O\neg a, Pc, O\neg(c \wedge a), \bigvee K$
$(\mathbf{D}_{-2}; \mathbf{D})$	p	T_p	$\bigvee K$
$(\mathbf{D}_{-2}; \mathbf{D})$	r	T_r^2	$\diamond(c \wedge a), O\neg a, Pc, O\neg(c \wedge a),$ $P(c \wedge m), \bigvee K$
$(\mathbf{D}_{-1}; \mathbf{D})$	p, r	T_p	$\bigvee K$

Table 2: Examples of defeasible inferences in the case of knifed murder, based on knowledge base $\{\diamond(c \wedge a), O\neg a, Pc\}$. We have $T_p = \bigcap_{\Gamma \in MC_{\mathbf{D}_{-1}}(K)} Cn_{\mathbf{D}_{-1}}(\Gamma)$, $T_r^1 = \bigcap R_{\mathbf{D}_{-2}; \mathbf{D}_{-1}}(K)$, and $T_r^2 = \bigcap R_{\mathbf{D}_{-2}; \mathbf{D}}(K)$.

8 Preferences on premises

This section describes further research on defeasible inferences defined by preferences on premises. Example 7.1 shows that the premise-based ordering equalises all inconsistent results, and then only provides the disjunction of all inconsistent formulas to receive a consistent conclusion. If we have different priorities on the premises, do we have different defeasible consequences? To answer this question, we define defeasible deontic logics based on different priorities on the premises. All defeasible deontic logics instantiated in this section are handled by a more general *undermining* mechanism.

Preferences over arguments can be distinguished according to different taxonomies of premises. Here we investigate two approaches. One suggests splitting arguments into two parts, such that some deontic formulas are more preferable than others. This provides us with different kinds of preferences over arguments based on

the language types of their premises. Another approach follows that discussed by Modgil and Prakken [65], and it divides arguments by strict and defeasible knowledge.

8.1 Preferences on language types

Now, we distinguish between arguments based on the different kinds of premises they have. We first categorise arguments by the premises in which we are interested—in the form of $\diamond\varphi$, $O\varphi$ or $P\varphi$ —and then we propose six different orderings according to these categories. Here, we categorise the deontic language by modalities. We say that φ is a non-permissible formula denoted as $\varphi \in \mathcal{L}_P^-$ iff there is no P -modality appearing in φ , that formula φ is a non-obligatory formula denoted as $\varphi \in \mathcal{L}_O^-$ iff there is no O -modality appearing in φ , that formula φ is a non-factual formula denoted as $\varphi \in \mathcal{L}_{\diamond}^-$ iff there is no \diamond - or \square -modality appearing in φ , that formula φ is a permissible formula denoted as $\varphi \in \mathcal{L}_P$ iff the only modality appearing in φ is P -modality, that formula φ is an obligatory formula denoted as $\varphi \in \mathcal{L}_O$ iff the only modality appearing in φ is O -modality, and that formula φ is a factual formula denoted as $\varphi \in \mathcal{L}_{\diamond}$ iff the only modalities appearing in φ are either \diamond - or \square -modality. For example, we have $O\neg(c \wedge a) \wedge Pm \in \mathcal{L}_{\diamond}^-$.

Definition 8.1 (Argument Properties, continued). Let $A, B \in \mathcal{A}$ be arguments and E a set of arguments. The partial order \leq is: strictly factual iff we have ($A \leq B$ iff $Prem(B) \subseteq \mathcal{L}_{\diamond}$), strictly obligated iff ($A \leq B$ iff $Prem(B) \subseteq \mathcal{L}_O$), strictly permitted iff ($A \leq B$ iff $Prem(B) \subseteq \mathcal{L}_P$), obligated iff ($A \leq B$ iff $Prem(B) \subseteq \mathcal{L}_P^-$), permitted iff ($A \leq B$ iff $Prem(B) \subseteq \mathcal{L}_O^-$), or deontic iff ($A \leq B$ iff $Prem(B) \subseteq \mathcal{L}_{\diamond}^-$).

We use \leq^{τ} to denote the τ -ordering with $\tau \in \{f, o^s, a^s, o, a, d\}$, where f stands for strictly factual, o^s for strictly obligated, a^s for strictly permitted, o for obligated, a for permitted, and d for deontic. We define the argumentation framework, defeat relation and different extensions as before.

We first consider the arguments with dominant premises and then the arguments with non-dominant ones. We define $K^{\tau} = \{B \in \mathcal{A} \mid \forall A \in \mathcal{A}(A \leq^{\tau} B)\}$ where \mathcal{A} is the set of arguments on the basis of AT with a knowledge base K . So, by a given K^{τ} , we only collect all the arguments that have their premises as obligatory, permitted, or other language types. Given such an \mathcal{A} , an argument A is either in K^{τ} or in $K - K^{\tau}$, but not in both. Accordingly, we can have the following proposition with regard to the orderings on premises, which provides a more general method for searching for stable extensions. Briefly speaking, we first deal with the dominant arguments based on K^{τ} as in Proposition 7.1.1, and then we deal with the non-dominant ones based on $K - K^{\tau}$. During this process, we need to ensure that each

new selected argument does not conflict with the old arguments.

Proposition 8.1. Consider the deontic language \mathcal{L} and a pair of two monotonic logics $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$. Let AF , corresponding to $\langle AT, \leq^\tau \rangle$, be an abstract argumentation framework $(\mathcal{A}, \mathcal{D})$ such that AT is based on $(\mathbf{S}^-; \mathbf{S}^+)$, K is a knowledge base, and $\tau \in \{f, o^s, a^s, o, a, d\}$. We construct a τ -premise set generated by K as $\bigcup_{n \in \mathbb{N}} E_n$ such that :

$$E_0 = \{D \in \mathcal{A} \mid F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma_1)\} \text{ for some } \Gamma_1 \in MC_{\mathbf{S}^+}(K^\tau)$$

$$E_{n+1} = \begin{cases} E_n \cup \{D \in \mathcal{A}\}, & \text{if } \exists \Gamma_2 \in MC_{\mathbf{S}^+}(K - K^\tau) \text{ such that} \\ & (i) F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma_2) \text{ and} \\ & (ii) F(D) \cup F(E_n) \text{ is } \mathbf{S}^+ \text{-consistent;} \\ E_n, & \text{otherwise.} \end{cases}$$

Then:

- E is a τ -premise set generated by K iff E is a stable extension regarding K .

The following proposition shows two connections of stable extensions from the preferences they are based on. We say a preference \leq^1 is less informative than another \leq^2 if and only if $K^{\leq^1} \subseteq K^{\leq^2}$. First, roughly speaking, a stable extension based on the less informative preference contains all the information in one stable extension based on the more informative preference. Secondly, the more informative preference constructs more stable extensions than the less informative one. All this shows that a more informative preference leads to a larger stable extension.

Proposition 8.2. Consider the deontic language \mathcal{L} and a combination of two monotonic logics $(\mathbf{S}^-; \mathbf{S}^+)$. Let AF_i , corresponding to $\langle AT, \leq^i \rangle$, be an abstract argumentation framework $(\mathcal{A}, \mathcal{D}_i)$ such that AT is based on $(\mathbf{S}^-; \mathbf{S}^+)$, K is a knowledge base, $i \in \{1, 2\}$, and \leq^i is a preference on premises. Let $Stable(AF_i)$ be the set of all stable extensions w.r.t. AF_i .

- If $K^{\leq^1} \subseteq K^{\leq^2}$, then $E \in Stable(AF_1)$ implies $\exists E' \in Stable(AF_2)$ s.t. $E' \subseteq E$.
- If $K^{\leq^1} \subseteq K^{\leq^2}$, then $|Stable(AF_1)| \leq |Stable(AF_2)|$.

This implies that $|Stable(\langle AT, \leq^{o^s} \rangle)| \leq |Stable(\langle AT, \leq^o \rangle)|$.

The following example illustrates the effects brought by the general undermining mechanism compared to those of rebuttal as proposed in Section 7.

Example 8.1 (Air Pollution, continued). Given the set $K = \{\diamond(c \wedge a), O \neg a, Pc\}$ as the premises again, we have arguments A, A', B, B', C, C' defined as follows:

- $A = O\neg a$
- $A' = A, C \Leftrightarrow \neg Pc$
- $B = Pc$
- $B' = B, C \Leftrightarrow \neg O\neg a$
- $C = \diamond(c \wedge a)$
- $C' = A, B \Leftrightarrow \neg\diamond(c \wedge a)$

The preferences over arguments A, A', B, B', C, C' are presented as follows:

- $f: C \geq A, A', B, B', C'$
- $o^s: A \geq A', B, B', C, C'$
- $a^s: B \geq B', A, A', C, C'$
- $o: A, A', C \geq B, B', C'$
- $a: B, B', C \geq A, A', C'$
- $d: A, B, C' \geq A', B', C$

The dominant arguments can defeat the non-dominant ones. See Figure 3 for an example of $(\mathbf{D}_{-2}; \mathbf{D})$. This strategy leads to the different defeasible consequences in Table 3. All these consequences are consistent in specific defeasible deontic logics, and explain why the chosen language types for arguments are better than the others. However, they provide different intuitive results compared to those of the running example. In other words, the common-sense reasoning of knifed murder does not follow the argumentation machinery developed with reference to the language type.

Table 3 presents another way of showing how the methodology of formal argumentation has an effect on the logical consequences of defeasible deontic logic. In the Introduction, we mentioned that one assumption behind the defeasibilities is that obligations defeat permissions. This assumption is, for instance, illustrated by the defeasible inferences $|\sim_{\mathbf{D}_{-2}; \mathbf{D}_{-1}}^{o^s \forall}$ and $|\sim_{\mathbf{D}_{-2}; \mathbf{D}}^{o \forall}$. Some other examples in Table 3 illustrate the other assumptions regarding the deontic modalities, such as permission defeating obligation with the defeasible inferences $|\sim_{\mathbf{D}_{-2}; \mathbf{D}}^{a^s \forall}$ and $|\sim_{\mathbf{D}_{-1}; \mathbf{D}}^{a \forall}$. In other words, assumptions of how one type of language defeats another explain these defeasible consequences.

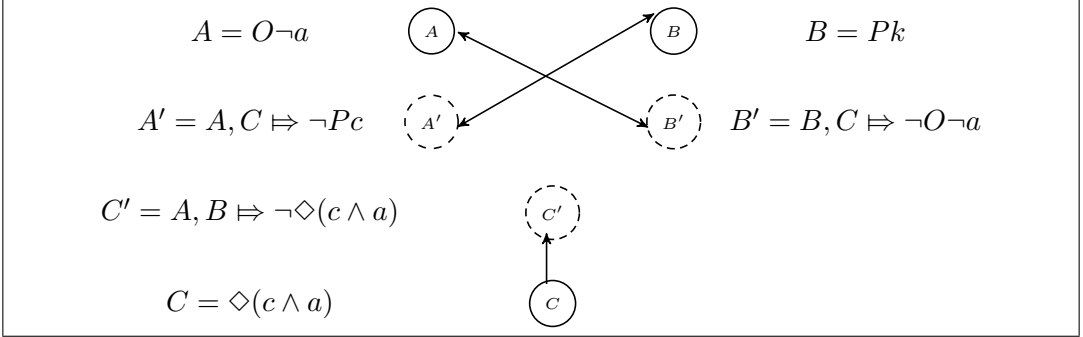


Figure 3: Some of the defeats among arguments A, A', B, B', C, C' based on $(\mathbf{D}_{-2}; \mathbf{D})$ in the strictly factual ordering \leq^f . Straight arrows are defeats among these arguments.

	Order	Example of Consequences $\{\diamond(c \wedge a), O\neg a, Pc\}$
$(\mathbf{D}_{-2}; \mathbf{D}_{-1})$	f	$\diamond(c \wedge a)$
$(\mathbf{D}_{-2}; \mathbf{D}_{-1})$	o^s	$O\neg a, O\neg(c \wedge a)$
$(\mathbf{D}_{-2}; \mathbf{D})$	a^s	$Pc, P(c \wedge a), P(c \wedge m)$
$(\mathbf{D}_{-2}; \mathbf{D})$	o	$\diamond(c \wedge a), O\neg a, O\neg(c \wedge a), \neg Pc, \neg P(c \wedge a)$
$(\mathbf{D}_{-1}; \mathbf{D})$	a	$\diamond(c \wedge a), Pc, \neg O\neg a, P(c \wedge a), P(c \wedge m)$
$(\mathbf{D}_{-1}; \mathbf{D})$	d	$O\neg a, Pc, \neg\diamond(c \wedge a), O\neg(c \wedge a), P(c \wedge a)$

Table 3: Various defeasible consequences of knowledge base $\{\diamond(c \wedge a), O\neg a, Pc\}$, depending on various preferences on language types

8.2 Preference on knowledge bases

Section 6 mainly focuses on the influence of preference based on *rules*, while Section 8.1 discusses reasoning on arguments based on *premises* regarding language type. Notice that the premises of arguments are generated from the knowledge base. This section moves towards a further step—considering the effect of preference on the knowledge base. In general, ASPIC⁺ takes two kinds of knowledge into consideration: defeasible and strict knowledge. The former takes the premises of arguments that are possible to defeat, while the latter takes the premises of arguments that cannot be defeated [65].

Definition 8.2 (Argument Properties, continued). Let $K = K_s \cup K_d$ be a knowledge base where K_s is called a set of strict knowledge and K_d is called a set of defeasible knowledge. Let A be an argument. Then A is firm iff $Prem(A) \subseteq K_s$, or plausible

iff $Prem(A) \cap K_d \neq \emptyset$. The partial order \leq^{fr} is firm iff ($A \leq B$ iff B is firm).

We then define $K^{fr} = \{B \in \mathcal{A} \mid \forall A \in \mathcal{A} (A \leq^{fr} B)\}$.

Proposition 8.3. Given the deontic language \mathcal{L} and a pair of two monotonic logics $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$, let AF , corresponding to $\langle AT, \leq \rangle$, be an abstract argumentation framework $(\mathcal{A}, \mathcal{D})$ such that AT is based on $(\mathbf{S}^-; \mathbf{S}^+)$, $K = K_s \cup K_d$ is a knowledge base, and \leq is firm. We construct a fr -premise set generated by K as $\bigcup_{n \in \mathbb{N}} E_n$ such that:

$$E_0 = \{D \in \mathcal{A} \mid F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma_1)\} \text{ for some } \Gamma_1 \in MC_{\mathbf{S}^+}(K^{fr})$$

$$E_{n+1} = \begin{cases} E_n \cup \{D \in \mathcal{A}\}, & \text{if } \exists \Gamma_2 \in MC_{\mathbf{S}^+}(K - K^{fr}) \text{ such that} \\ & (i) F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma_2), \text{ and} \\ & (ii) F(D) \cup F(E_n) \text{ is } \mathbf{S}^+ \text{-consistent;} \\ E_n, & \text{otherwise.} \end{cases}$$

Then:

- E is a fr -premise set generated by K iff E is a stable extension regarding K .

The defeasible inference defined below follows the sceptical account of defeasible reasoning [48]. In argumentation theory, these *sceptical* inferential consequences result from the arguments contained in the intersection of all stable extensions.

Definition 8.3 (Defeasible Inferences). Let $\Gamma \subseteq \mathcal{L}$ and let $\phi \in \mathcal{L}$. Let $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics, and let \leq^τ be a τ -ordering such that $\tau \in \{f, o^s, a^s, o, a, d, fr\}$. Let AT be the Γ -argumentation theory based on $(\mathbf{S}^-; \mathbf{S}^+)$ iff the argumentation theory AT obtains with $K = \Gamma$, and let $AF^\tau = \langle AT, \leq^\tau \rangle$. The defeasible inference $\|\sim_{\mathbf{S}^-; \mathbf{S}^+}^\tau$ is defined as follows:

- $\Gamma \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \phi$ iff every stable extension of the Γ - AT based on $(\mathbf{S}^-; \mathbf{S}^+)$ corresponding to AF^τ contains an argument A with $Conc(A) = \phi$.

We define the closure operator corresponding to this inference relation as usual: $\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma) = \{\phi \mid \Gamma \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \phi\}$. Moreover, we write $\|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \phi$ when $\emptyset \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \phi$.

The proposition below presents a uniform way to construct all defeasible inference relations for all these preferences on *premises* rather than *rules* based on the result of Proposition 8.3.

Proposition 8.4. Let $\Gamma \subseteq \mathcal{L}$, $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics, let \leq^τ be a τ -ordering with $\tau \in \{p, f, o^s, a^s, o, a, d, fr\}$, and let K be a knowledge base of AT . We define

- a P-set generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$ as $\bigcup_{n \in \mathbb{N}} P_n$, such that

$$P_0 = Cn_{\mathbf{S}^+}(\Gamma)$$

$$P_{n+1} = \begin{cases} P_n \cup \{\varphi\}, & \text{if } \exists \Gamma' \in MC_{\mathbf{S}^+}(K - K^\tau) \text{ such that} \\ & (i) \varphi \in Cn_{\mathbf{S}^+}(\Gamma') \text{ and} \\ & (ii) \{\varphi\} \cup P_n \text{ is } \mathbf{S}^+\text{-consistent;} \\ P_n, & \text{otherwise;} \end{cases}$$

where $\Gamma \in MC_{\mathbf{S}^+}(K^\tau)$.

The P-collection $P_{\mathbf{S}^-; \mathbf{S}^+}(K)$ generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$ is the set of all P-sets generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$. Then,

- $C_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(K) = \bigcap P_{\mathbf{S}^-; \mathbf{S}^+}(K)$.

Corollary 8.1. Proposition 7.2.1 is a special case of Proposition 8.4.

Now, we check the defeasible inferences based on the division of strict and defeasible knowledge in the mixed mechanism. We study how the argumentation machinery regarding the impact of this kind of category on a knowledge base helps to explain the results.

Example 8.2 (Air Pollution, continued). We consider the defeasible consequences $C_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(K_i)$ shown in Table 4, given the following divisions of knowledge bases K_i where $1 \leq i \leq 4$:

- $K_1 = K_s \cup K_d$ where $K_s = \{\diamond(c \wedge a)\}$ and $K_d = \{O\neg a, Pc\}$
- $K_2 = K_s \cup K_d$ where $K_s = \{\diamond(c \wedge a), O\neg a\}$ and $K_d = \{Pc\}$
- $K_3 = K_s \cup K_d$ where $K_s = \{\diamond(c \wedge a), Pc\}$ and $K_d = \{O\neg a\}$
- $K_4 = K_s \cup K_d$ where $K_s = \{\diamond(c \wedge a), O\neg a, Pc\}$ and $K_d = \emptyset$

Both the defeasible consequences based on the different language types and those based on the distinction between strict and defeasible knowledge only partially capture the intuition of knifed murder. See the latter case in Table 4.

The defeasible consequences in Table 4 reflect the idea in argumentation that an argument with strict knowledge defeats those with defeasible knowledge. In particular, the results of $C_{\mathbf{D}_{-2}; \mathbf{D}}^{fr}(K_2)$ illustrate how permissions are defeated by obligations generally, while the results of $C_{\mathbf{D}_{-2}; \mathbf{D}}^{fr}(K_3)$ show how permissions defeat obligations. See Figure 4 for an example of how K_2 generates the conclusions based on $(\mathbf{D}_{-2}; \mathbf{D})$ in the firm ordering of \leq^{fr} .

$(\mathbf{S}^-; \mathbf{S}^+)$	Order	K_i	$\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^\tau(K_i)$
$(\mathbf{D}_{-2}; \mathbf{D}_{-1})$	<i>fr</i>	K_1	$\diamond(c \wedge a)$
$(\mathbf{D}_{-2}; \mathbf{D})$	<i>fr</i>	K_1	$\diamond(c \wedge a)$
$(\mathbf{D}_{-1}; \mathbf{D})$	<i>fr</i>	K_1	$\diamond(c \wedge a)$
$(\mathbf{D}_{-2}; \mathbf{D})$	<i>fr</i>	K_2	$\diamond(c \wedge a), O\neg a, O\neg(c \wedge a), \neg Pc, \neg P(c \wedge a)$
$(\mathbf{D}_{-2}; \mathbf{D})$	<i>fr</i>	K_3	$\diamond(c \wedge a), \neg O\neg a, \neg O\neg(c \wedge a), Pc, P(c \wedge a), P(c \wedge m)$
$(\mathbf{D}_{-1}; \mathbf{D})$	<i>fr</i>	K_3	$\diamond(c \wedge a), \neg O\neg a, \neg O\neg(c \wedge a), Pc, P(c \wedge a), P(c \wedge m)$
$(\mathbf{D}_{-2}; \mathbf{D}_{-1})$	<i>fr</i>	K_4	$\bigvee K_4$
$(\mathbf{D}_{-2}; \mathbf{D})$	<i>fr</i>	K_4	$\bigvee K_4$

Table 4: Defeasible consequences based on a knowledge base K_i

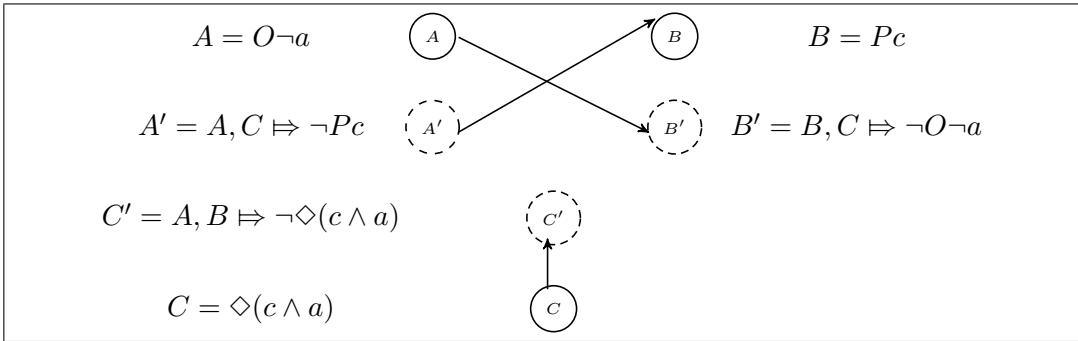


Figure 4: Some of the defeats among arguments A, A', B, B', C, C' based on K_2 and $(\mathbf{D}_{-2}; \mathbf{D})$ in the firm ordering \leq^{fr} . Straight arrows are defeats among these arguments.

9 Supra-classical inference

In the previous section, we presented various possible ways to define defeasible consequences. It can be defined according to preferences over *rules* (c.f. Section 6). It can also rely on preference orders over *premises*, like what we defined in Section 8. These variants define defeasible consequences depending on the ways in which different information in the knowledge base are considered to be more or less preferable: whether the formulas of the premises are obligated, permitted, or factual; or whether they are classed as strict or defeasible knowledge. We think they are good strategies for modelling defeasible consequences. They assume different assumptions behind defeasibilities regarding the structures of deontic modalities. In Section 7 and Section 8, we defined so-called sceptical inferences based on stable extensions. We could also consider the architectures of argumentation semantics, for instance credulous

inference [49], which is another common inference in non-monotonic reasoning. We leave that discussion to future work. We call all the defeasible inferences defined previously *supra-classical* inferences [62] because they provide more information than classical inferences.²

In this section, we check relations between supra-classical inferences and monotonic inferences. We define the subsystems in this way: $\mathbf{S}' \subseteq \mathbf{S}$, i.e. the theorems in subsystem \mathbf{S}' are contained in its extension \mathbf{S} . And $\Gamma \vdash_{\mathbf{S}} \varphi$ is defined in Definition 5.5, representing derivations from premises Γ to conclusion φ . We first have the following proposition regarding atomic propositions:

Proposition 9.1. Let $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics. Now we have $p \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau} p$ but $\{p, \neg p\} \not\vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau} p$, with $\tau \in \{p, r, f, o^s, a^s, o, a, d, fr\}$.

The supra-classical inferences we defined are non-monotonic. The following proposition offers a general result on their connections.

Proposition 9.2. Let $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics. We have the following relation regarding supra-classicality:

$$\vdash_{\mathbf{S}^- \subseteq} \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau} \subseteq \vdash_{\mathbf{S}^+}$$

where $\tau \in \{p, r, f, o^s, a^s, o, a, d, fr\}$.

Now, we shall evaluate all the defeasible deontic logics defined in this article. First, we consider whether the extensions instantiated satisfy the rationality postulates. The main tool for studying formal argumentation in the setting of ASPIC⁺ is based on using rationality postulates [16]. It immediately follows from Propositions 7.1, 8.1 and 8.3, that all the rationality postulates are satisfied. This can also be proven as a corollary of the more general theorems of Caminada [16] and those of Modgil and Prakken [65].

Another evaluation we consider is that of a summary of the logical properties of all defeasible deontic logics. The following proposition shows whether these defeasible deontic logics satisfy some non-monotonic properties, which are the *rationality postulates* mentioned previously.

Proposition 9.3. Given $\tau \in \{p, r, f, o^s, a^s, o, a, d, fr\}$ as one of the preferences defined and $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ as a pair of two monotonic

²It is also possible to define supra-classical inference in accordance with the ‘classical’ state of the art. For instance, recent work on substructural deontic logics [24, 38] studied several ways to exclude the undesired classical inferential patterns while still trying to maintain a certain amount of restricted monotonicity in control of different substructural rules for modal operators.

logics, we will check whether the defeasible deontic logics defined in this article satisfy the following standard properties regarding non-monotonicity, where we simplify $\|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}$ to \Vdash :

1. Reflexivity: $\Gamma \Vdash \varphi$ where $\varphi \in \Gamma$
2. Cut: if $\Gamma \cup \{\psi\} \Vdash \chi$ and $\Gamma \Vdash \psi$, then $\Gamma \Vdash \chi$
3. Cautious Monotony: if $\Gamma \Vdash \psi$ and $\Gamma \Vdash \chi$, then $\Gamma \cup \{\psi\} \Vdash \chi$
4. Left Logical Equivalence: if $Cn_{\mathbf{S}^+}(\Gamma) = Cn_{\mathbf{S}^+}(\Gamma')$ and $\Gamma \Vdash \chi$, then $\Gamma' \Vdash \chi$
5. Right Weakening: if $\vdash_{\mathbf{S}^+} \varphi \rightarrow \psi$ and $\Gamma \Vdash \varphi$, then $\Gamma \Vdash \psi$
6. OR: if $\Gamma \Vdash \varphi$ and $\Gamma' \Vdash \varphi$, then $\Gamma \cup \Gamma' \Vdash \varphi$
7. AND: if $\Gamma \Vdash \psi$ and $\Gamma \Vdash \chi$, then $\Gamma \Vdash \psi \wedge \chi$
8. Rational Monotony: if $\Gamma \Vdash \chi$ and $\Gamma \not\vdash \neg\psi$, then $\Gamma \cup \{\psi\} \Vdash \chi$

The results are shown in Table 8.

Properties	$\ \sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}$	$\ \sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}$
Reflexivity	\checkmark^*	No
Cut	\checkmark	\checkmark
Cautious Monotony	\checkmark	\checkmark
Left Logical Equivalence	\checkmark	\checkmark
Right Weakening	No	\checkmark
OR	No	No
AND	\checkmark	\checkmark
Rational Monotony	\checkmark	\checkmark

Table 5: This is a summary regarding various principles of defeasibilities. Notice that $\tau \in \{p, f, o^s, a^s, o, a, d, fr\}$. The symbol \checkmark^* indicates that this property is satisfied when the given knowledge base is consistent in \mathbf{S}^- .

Now, we provide some counterexamples to the non-monotonic properties.

Example 9.1 (Invalidities of Reflexivity). Let $\tau \in \{p, f, o^s, a^s, o, a, d\}$. By Examples 7.1, 8.1 and 8.2, we know that $\|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}$ is not reflexive.

Example 9.2 (Invalidities of Right Weakening). Let $\Gamma = \{O\neg a, \diamond(c \wedge a), Pc\}$ be a sentence illustrating the knifed murder scenario. Then,

- $\Gamma \Vdash_{\mathbf{D}_{-2};\mathbf{D}}^{r\forall} Pc$ but $\Gamma \not\vdash_{\mathbf{D}_{-2};\mathbf{D}}^{r\forall} P(c \wedge a)$.

Example 9.3 (Invalidities of OR). Let $\Gamma = \{\diamond(c \wedge a), Pc\}$ and $\Gamma' = \{O\neg a, Pc\}$. Then,

- $\Gamma \Vdash_{\mathbf{D}_{-2};\mathbf{D}}^{r\forall} P(c \wedge a)$ and $\Gamma' \Vdash_{\mathbf{D}_{-2};\mathbf{D}}^{r\forall} P(c \wedge a)$ but not $\Gamma \cup \Gamma' \Vdash_{\mathbf{D}_{-2};\mathbf{D}}^{r\forall} P(c \wedge a)$.

10 Extending to various modal languages

In this section, we extend our discussion from monadic to conditional permission and obligation with different modal languages. We can extend the language into conditional obligation and permission, and then explore *deontic detachment* and *contrary-to-duty reasoning* [47, 59]. This deontic problem was originally phrased by Chisholm in [18]. We will discuss one variant of contrary-to-duty reasoning in Section 10.1. We will explore obligation and permission in legal reasoning in Section 10.2. Rights and duties are important concepts in Hohfeld’s theory of legal rights [46]. In particular, legal power and liability involve the notions of agency and actions, thereby capturing another dimension of legal rights. We need a modal language to express these concepts. The third approach we will explore focuses on permission to know [6], in particular, the right to know as described in Section 10.3. Finally, we discuss permissive norms in Section 10.4.

10.1 Conditional permission

One important issue with conditional obligation, or prima facie obligation, is the problem of deontic detachment. It was first discussed by Chisholm [18] and was later known as “Chisholm’s paradox”, or “contrary-do-duty paradox” [74]. As widely agreed, from an intuitive point of view, given a set of statements of conditional obligations like those in Example 10.1, the paradox is consistent and all its members are logically independent of one other. The challenge is that when formalising these statements, it turns out that they are neither logically consistent nor independent. This challenge can be addressed by the following variant.

Example 10.1 (Deontic Detachment: Obligation). Intuitively speaking, the following set of sentences are consistent, and their members are logically independent of one other.

- (A) It ought to be the case that Jones does not eat fast food for dinner.

- (B) It ought to be the case that if Jones does not eat fast food for dinner, then he does not go to McDonald's.
- (C) If Jones eats fast food for dinner, then he ought to go to McDonald's.
- (D) Jones eats fast food for dinner.

A conditional obligation “It ought to be the case that if ψ then φ ” can be represented by the formula $O(\varphi \mid \psi)$. Then, an unconditional obligation $O\varphi$ is stipulated as an abbreviation of $O(\varphi \mid \top)$. Then, the above-mentioned sentences are formulised as:

- (A') $O\neg f$
- (B') $O(\neg g \mid \neg f)$
- (C') $f \rightarrow Og$
- (D') f

where f is short for “Jones eats fast food for dinner” and g is short for “Jones goes to McDonald's”. By the following CTO pattern for cumulative transitivity

$$\text{CTO: } \frac{O(\varphi \mid \psi \wedge \chi), O(\psi \mid \chi)}{O(\varphi \mid \chi)}$$

we then have conclusion $O\neg g$ from (A') and (B') as well as conclusion Og from (C') and (D'). As the argument in Example 1.1 shows, the results turn out to be inconsistent.

Many proposals have been suggested to solve this problem. One possible way is to interpret the conditionals as anankastic conditionals [20], also known as hypothetical imperatives. Another possible way is to adopt different kinds of non-monotonic tools for the representation and reasoning [74, 59, 86]. There is a consensus on deontic detachment that techniques from non-monotonic reasoning can be used to handle reasoning of prima facie obligation. However, there is less consensus about how these techniques can be used to deal with prima facie obligation. Please refer to the recent review by Pigozzi and van der Torre [71] for details. We only present the key idea here in order to introduce a problem regarding prima facie permission.

The variant presented below illustrates a scenario of deontic detachment regarding prima facie permission.

Example 10.2 (Deontic Detachment: Permission). The following scenario is a variant of an example regarding permission by Prakken and Sergot [74]. It contains statements that are consistent and intuitive in natural language:

- (A) A dog is permitted if it is a guide dog for a blind man.

- (**B**) It is permitted that if there is a dog, then there is a fence and it is painted white.
- (**C**) It ought to be that there is no fence.
- (**D**) It is possible that there is a fence and it is painted white.

Now, a conditional permission “It is permitted that if ψ then φ ” is formulated as $P(\varphi \mid \psi)$, and then the unconditional version $P\varphi$ is short for $P(\varphi \mid \top)$. These four statements can be represented as follows:

- (**A'**) $P(d \wedge g)$
- (**B'**) $P(w \wedge f \mid d)$
- (**C'**) $O\neg f$
- (**D'**) $\diamond(w \wedge f)$

where d stands for “There is a dog”, g for “It is a guide dog for a blind man”, w for “It is painted white”, and f for “There is a fence”. We now consider two possible patterns of prima facie permission. The first one is the so-called strengthening antecedent, denoted as SA and presented as follows:

$$\text{SA: } \frac{P(\varphi \mid \psi), \Box(\chi \rightarrow \psi)}{P(\varphi \mid \chi)}$$

The second CTP pattern we consider is used as cumulative transitivity:

$$\text{CTP: } \frac{P(\varphi \mid \psi \wedge \chi), P(\psi \mid \chi)}{P(\varphi \mid \chi)}$$

There are two issues regarding the reasoning of prima facie permission. First, by using the SA pattern, we then infer (**B''**) $P(w \wedge f \mid d \wedge g)$ from (**B'**). And then by using CTP, we get (**C''**) $P(w \wedge f)$ from (**B''**) and (**A'**). Until now, we have not used the FCP pattern or the OWP pattern in the **D** system. Yet, we have already reached an uneasy situation according to our intuition. It is permitted to have a white fence no matter which precondition is given (i.e. $P(w \wedge f)$). Why is this situation uneasy? From (**C'**), we usually infer $O\neg(w \wedge f)$ by using K_O and NEC_O . Further, when taking this (**C''**) with the assumption (**C'**) by applying OWP, it implies something contrary to (**D'**). In other words, their results are inconsistent.

These two issues lead to two questions. The first question is which axioms or rules are appropriate for developing a defeasible logic of conditional permission? The second question is about preferences. What kind of preferences on arguments

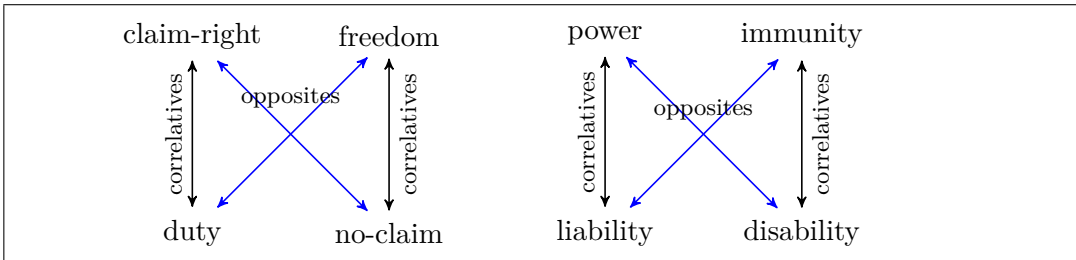


Figure 5: Hohfeldian right relations [Sergot, 2013]

are more useful for distinguishing a reasonable prima facie permission from the others? Our ASPIC⁺-based defeasible logic may help to handle the problems identified above. We leave the issue of how to apply our ASPIC⁺-based defeasible logic to handling the problem of prima facie permission to future work.

10.2 Rights and duties

In legal reasoning, the terms rights and duties are more significantly used than permissions and obligations. It is necessary to consider agency when defining these legal notions. In this section, we review the discussion of rights and duties in legal and deontic literature [46, 79]. We propose a basic representation of rights and duties by using additional modal operators for agency and actions. Various legal rights can be composed in our modal language, for instance, the right to privacy as well as active and passive rights.

Rights play a central role in deontic logic, since they point to a crucial social phenomenon: how agents' social or *normative positions* depend on others and on others' positions. It is a well-known fact that talking about "rights" in itself is ambiguous. This ambiguity easily leads to conceptual obscurity, and so a hundred years ago, an American legal theorist differentiated between various possible meanings of legal rights [46, 64]. In Hohfeld's system, which consists of four different right relations, there are four different rights, and each type of right matches a given type of duty on the other side: someone's right always means *someone else* has a duty. Sergot [79] calls these pairs of rights, or *normative positions*, correlative relations, as they always come together. From a logical point of view, the rights with the same correlative relation will be equivalent. The system Hohfeld designed can be graphically represented with rights in the upper row and duties in the lower row (see Figure 5). The opposite relations show the effect of a negation of someone's rights on another's duty and vice versa. More details will be shown in the next paragraphs.

These positions are most apparent in the case of a contract of sale. A seller's

claim-right to the purchase price obviously means the buyer’s duty to give her the money. But this phenomenon is far more general. Regarding epistemic positions, if an agent has a claim-right to know something, it means that another agent has a duty *directed towards* the previous agent to tell him: $R_a[b]K_a\varphi \Leftrightarrow O_{b,a}[b]K_a\varphi$, where $R_a\varphi$ is read as “Agent a has a right that φ ”, $[a]\varphi$ is read as “Agent a executes an action to make φ true”, $K_a\varphi$ is read as “Agent a knows that φ ”, and directed obligation $O_{b,a}\varphi$ is read as “Agent b has a duty towards agent a that φ ”. However, if an agent has the freedom to know something, that only means that the other agent has no claim-right towards him that he doesn’t get to know. We usually only consider this position as real freedom to know when no other agent has a claim-right that the previous agent shouldn’t get to know that thing. We usually refer to such a position as *permission*, which we will analyse in Section 10.3.

The square on the right-hand side exhibits a very similar structure. However, those positions are dynamic [53, 58] and are about the *potential* to change others’ rights and duties [64]. For example, if we consider the right to know something as a power, that means that the agent having this power can impose a duty directed towards another agent—whose position is called liability in this system—to let her know: $Power_{a,b}[a]O_{b,a}[b]K_a\varphi$. Meanwhile, if someone has an immunity regarding her knowledge, that would mean that the other agent has no power to impose a duty on her to tell him. For details on formalising power and the related positions, please refer to recent discussions by Dong and Roy [28, 29] and Markovich [64].

The agency of actions and the normative positions interact [79] in the sense that we can have freedoms regarding actions, so-called active rights, while we can only have claim-rights regarding other agents’ actions, i.e. passive rights. Power is active. It’s about an agent’s potential for action, commanding that the other agent does something in particular. In contrast, immunity is passive, because it is about the other agent’s lack of potential to do or demand something. Formalising the rights expressed in natural language is already very decisive because it will expose whose action should concern us; though it is not always clear what that action is [46]. As we see in the scope of a notion incorporating different things, there are different atomic rights pertaining to Hohfeldian types. The right to privacy covers many things, for example claim-rights towards everyone else to keep away from one’s private zone, that is, not to get to know about one’s private life. That means a long list of prohibitions: any action that might end up gathering and disclosing information pertaining to the private zone, as shown in Example 10.3.

These rights, like most rights apart from very few exceptions in modern constitutional democracies, are defeasible. This property means that their existence, the truth of the sentences expressing them, or the inferences we would draw from them, depend on the circumstances. It might happen that information subject to

someone's right to know is considered to be a secret for some reason. In that case, there would be a stronger argument for keeping it private. Although it would go against the agent's right to know the information, there would be no obligation to let that agent know it. Or it can be the other way around. Someone's right to privacy, which would mean that others shouldn't know about her private life and information, can also be defeasible. It may be the case that what that individual does in her private life could be dangerous to others, and the public interest is often a strong argument against the interests of the individual.

Sometimes, it is far from easy to decide which argument is stronger, the right to know or the right to privacy, and then we face a dilemma. In this situation, both cases can be represented using formal argumentation, which may lead to a precise decision.

10.3 Permission to know

Defeasible deontic logic formalises the practical reasoning of intelligent autonomous agents in situations involving uncertainty, conflicts and exceptions. In this section, we explain how to extend defeasible deontic logic with modal operators for epistemic notions like knowledge and belief. In particular, we use techniques from formal argumentation to represent common-sense reasoning to handle permission to know.

Example 10.3 provides a case study on representing claims regarding the right to know and permission to know. To do this, our modal language will be extended with the modalities K_i for knowledge and $\langle i \rangle$ for agent i 's ability to "see to it that", both indexed with agent i .

Example 10.3 (Sensitive data scenario). Anyone's health data counts as sensitive data and as such is subject to strong protection principles in most countries (in the European Union, there is the General Data Protection Regulation (the so-called GDPR) having the long title "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC"). This means that others are not allowed to know the data. However, if someone is ill and in need of medical treatment, we would all agree that doctors have to provide this medical treatment. But fulfilling this obligation requires that they know the health data of the agent (in this case, the patient).

Various claims arising from the example scenario above are visualised in Figure 6 below, together with a formalisation in ASPIC⁺. Moreover, the claims in the figure are grouped into two camps of arguments by vertical arrows. The four claims on

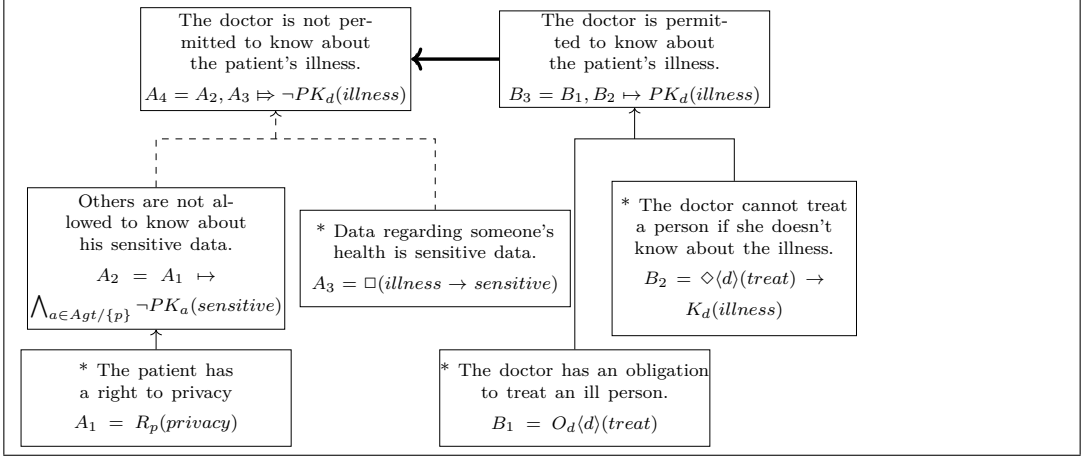


Figure 6: Two camps of arguments in the sensitive data scenario, each containing four claims. Each block is a claim containing the conclusion of an argument. The blocks marked with * are premises from the knowledge base.

the left constitute the argument that the doctor is not permitted to know the sensitive information, and the three claims to the right constitute the argument that the doctor is permitted to know that information. The top claims are the conclusions of their arguments, and the other claims support those conclusions. The fact that the doctor's permission to know prevails is modelled by an arrow from the top right box to the left top box. In this section, we consider only the natural language statements and their formalisation as argumentation elements. Employing the previous techniques we have developed, we can put forward a defeasible epistemic deontic logic for permission to know. We only introduce the basic idea here and leave further development to future research.

10.4 Permissive norms

Given the defeasibility of legal reasoning, many people stop using standard deontic logic and develop new rule-based systems instead. A drawback of this approach is the resulting gap between standard deontic logic and defeasible deontic logics. In this article, we study an alternative solution that builds a bridge from standard deontic logic to defeasible deontic logic.

Deontic logics range from the monadic modal logic of obligations and permissions, via the dyadic modal logic of conditionals, to rule-based systems for norms. Argumentation can be used at all three levels. A recent chapter in the *Handbook of Normative Multiagent Systems (NorMAS)* [da Costa Pereira *et al.*, 2018] focuses

on the most complex level of rules or norms. In this article, we focus more on the basic level of obligations and permissions, but we also give pointers on using argumentation for norms.

The main contribution of this article is the development of a rule-based system for monadic modal logic, and we propose that one natural way to apply our method to rule-based systems for norms is to handle *prima facie* obligations and permissions. The discussion in Section 10.1 sets a good direction for future research. However, to distinguish norms from obligations and permissions and then study norm compliance and violation, we may need to differentiate between defeasibility among rules and defeasibility among premises. That would require us to build a new architecture for a more complex inferential relation than what we have now. In that case, a hypothetical style of inference [61] would be worth considering. By doing so, we might not only provide a way to resolve conflicting norms and check compliance, but also study graded norms [23] in a quantitative setting. Then, it would be possible to interpret legal norms according to the presumptions assumed in particular contexts [61].

11 Related work

There are many existing approaches to defeasible deontic logics, including input/output logic [59, 67], the logic of imperatives [41, 42], paraconsistent deontic logic [21, 12], conditional deontic logic [79, 24], non-normal deontic logic (including ‘Seeing To It That’ (STIT) logic [50, 15] and neighbourhood semantics [3]), default logic [49], deontic preference logic [43], and dynamic deontic logic [82, 28]. They emphasise different perspectives on handling inconsistency with obligations, permissions, and many other aspects of norms. We categorise this research into two groups: those that handle norms on a propositional level, and those that study norms based on the structures of their deontic modalities. The summary is shown in Table 6.

Many defeasible deontic logics investigate obligation, prohibition, and permission through their propositional components. Input/output logic [59, 67] is one main approach. It proposes studying the different structures of dependency between input and output in a normative propositional system, and then defining, for instance, different kinds of obligations in terms of their specific structured outputs. When constructing the normative code, it is possible to take preferences into consideration [67]. The resulting logical consequences of input/output logic satisfy the defeasibility requirement, and thus provide more results than classical logics. As such, their consequences are *supra-classical* relations. The logic of imperatives [41, 42] and default logic [49] also have a similar spirit to handling norms: staying propo-

sitional, making it possible to have a preference, focusing on differently structured dependencies, and providing more consequences than classical logics. The logic of imperatives [41, 42], however, has axiomatisations in modal language. On the other hand, conditional deontic logics [79, 24] also treat norms on a propositional level, but usually their logical consequences are weaker or fewer than those of classical logics. These kinds of consequences are *infra-classical* relations.

Another key approach mainly considers norms on a modal level. Deontic preference logic [43], dynamic deontic logic [82, 28], and STIT logic [50, 15] are very famous modal frameworks in the literature. For instance, deontic preference logic [43] and dynamic deontic logic [82, 28] adopt Kripke models to capture preferential orders and then define modalities for obligation, prohibition, and permission. The preferences can be given [28] or derived [43, 82]. To follow the character of non-monotonicity, these logics try to keep the derived consequences as much as possible while excluding inconsistency in normative reasoning. So they generate supra-classical relations. There is an exception. For instance, when adopting a non-normal Kripke framework like STIT logic [50, 15] or neighbourhood semantics [3] to model norms, the logical consequences are fewer than usual. These infra-classical relations are the result of trade-offs in an attempt to balance the generality of their derivation results and their capacity to resolve a moral dilemma effectively [15].

Paraconsistent logic [21, 12] stands in between these two approaches. It usually handles norms at the propositional level, but still offers axiomatisation in the modal language [12]. Paraconsistent logic mostly concentrates on the dependent relations between the normative system and the results it leads to. Although the logical consequences in the early work of paraconsistent logic [21] are infra-classical, the most recent work [12] results in many more consequences, which we then call supra-classical.

Apart from these, adaptive logic [8, 9, 11], a currently active approach to defeasible reasoning, provides a set-theoretical configuration, similar to our developments of logical systems based on ASPIC⁺, to have a number of consistent sets derived from an inconsistent set if at all possible. In this logic, as with our approach, by having a lower-limit logic, each derivable formula is required to be consistent with those from the previous stage, given that abnormalities cannot be present. Two strategies, reliability and minimal abnormality, are used to develop sceptical or credulous inferences like our \forall - and \exists -types of inference. The key characteristic of adaptive logic is the way it interprets abnormality. In a “flat” adaptive logic, all the premises are equally preferred, while in a “prioritised” adaptive logic, premises are ordered in different layers [9]. However, all priorities are premise-based [9]. It is not clear how rule-based priorities can be captured in an adaptive framework.

Defeasible deontic logic [66] is a widely studied approach to normative reasoning

and offers a lot of formal tools in non-monotonic reasoning. Its main idea is to define defeasibility either in terms of consistency, governed under a set of formulas combined with a set of inference rules [34, 80, 37], or by providing a priority mechanism for overtaking less normal conclusions [49, 35]. For instance, Goble [34] provides an adaptive logic for handling different kinds of normative conflicts via the notion of abnormality. A formula is true from a set of formulas if and only if this formula is satisfied at every reliable and normal model. This inference relation highly depends on the sets of abnormalities and inferential rules on them. Straßer [80] follows Goble’s work and investigates dynamics in adaptive reasoning, while Governatori [37] proposes that multi-layered consistency for conditional obligations is captured by sequential operators for computing norms and their violations. In contrast, Horty [49] and Governatori [35] define defeasible consequences in terms of priorities over default rules. They both define priorities among default rules rather than over the arguments. Riveret *et al.* [76] propose a rule-based argumentation framework for representing conditional norms.

In a similar fashion, in order to be non-monotonic, facts in deontic update semantics [85, 86, 87] are updates that restrict the domain of the model. They make a fact ‘settled’ in the sense that it will never change again even after future updates of the same sort. Van Benthem *et al.* [82] use dynamic logic to place such a dynamic approach within standard modal logic. Dynamic logic includes reduction axioms and standard model theory. They rehabilitate classical modal logic as a legitimate tool to do deontic logic, and position deontic logic within the growing dynamic logic literature [28, 29]. In contrast to this dynamic approach, a recent work has been developed with weighted deontic modalities [25] in order to capture the ability of agents to make rational choices. Governatori *et al.* [39] have developed a possible world semantics for defeasible normative reasoning.

Connecting formal argumentation to deontic logic has been an increasingly active area of research in recent years [70]. An approach that is closely related to this article is called *logic-based instantiations of an argumentation framework* and can be traced back to the work of Benferhat *et al.* [13] and Cayrol [17]. Two key ideas highly related to this article were developed: Benferhat *et al.* [13] suggested the methodology of handling preferences in Dung-style argumentation theory via the concept of “level of paraconsistency”, while Cayrol [17] provided a more concrete method: investigating the link between stable extension and Maximally consistent sets (MCS) based on classical logic. Recent studies focus on connections between logic and argumentation, including checking the application of Gentzen proof theory on formal argumentation, as proposed by Arieli *et al.* [4], and instantiating ASPIC⁺ based on deontic modal logics about obligation and permission but for complete and grounded extension, as proposed by Beirlaen *et al.* [10]. A recent work that is

	based on a propositional level		
Properties	IOL	LI	PDL
Propositional Level	[59, 67]	[41, 42]	[21]
Modal Axiomatization		[41, 42]	[12]
With Given Preference	[67]	[42]	
With Derived Preference			
Dependency	[59, 67]	[41, 42]	[12]
Infra-Classicality			[21]
Supra-Classicality	[59, 67]	[41, 42]	[12]

Table 6: This is a summary of various frameworks of handling norms. IOL is short for input/output logic, LI is short for logic of imperatives, PDL is short for paraconsistent deontic logic, CDL is short for conditional deontic logic, STIT is short for STIT-logic, NS is short for neighbourhood semantics, DL is short for default logic, DPL is short for deontic preference logic, and DYDL is short for dynamic deontic logic.

close to our work, by Straßer and Arieli [81], presents an argumentative approach to normative reasoning using standard deontic logic as base logic. Similarly related, Liao *et al.* [57] represent three logics of prioritised norms by using argumentation. In addition, Glavaničová [33] studies how to let the logical principle of free choice permission be defeasible in non-monotonic adaptive logic. In contrast, Governatori *et al.* [36] provide a defeasible logic for computing strong and weak permissions, while Lam *et al.* [55] have developed a connection between ASPIC⁺ and defeasible logic. Dong *et al.* [27] have identified a possible way to develop AI logic for social reasoning with this ASPIC⁺-based method.

based on their deontic modalities						
CDL	STIT	NS	DL	DPL	DYDL	
[79, 24]			[49]			
	[50, 15]	[3]		[43]	[82, 28, 25]	
					[28]	
				[43]	[82]	
	[50, 15]		[49]			
[79, 24]	[50, 15]	[3]				
			[49]	[43]	[82, 28, 25]	

Table 7: Summary of various frameworks of handling norms, part 2.

12 Conclusions and future work

In this article, ASPIC⁺ connects formal argumentation to non-monotonic logic. We believe this approach benefits both areas. For formal argumentation, the resulting non-monotonic logics can be studied to provide new insights into the argumentation systems adopted, for example we can apply the logical results of our ASPIC⁺-logic to learn more about the effect of the argumentation semantics adopted. For non-monotonic logics, the underlying argumentation theory can be used to explain deontic conclusions. Our case study on using the logic of obligations and permissions provides first evidence of this.

Within this general ambitious setting, the contributions of this article are as follows. First, with regard to the definitions, in Definitions 5.6 and 5.7 we show how to use two logics in ASPIC⁺, and in Definitions 7.1 and 8.3 we show how to build a defeasible modal logic on top of ASPIC⁺. With regard to formal results, Proposition 7.2 and 8.4 characterise the consequences of our defeasible deontic logics. As these representation theorems show, our defeasible deontic logics can be built without ASPIC⁺. The role of ASPIC⁺ is likely to be an interpreter. It provides an intuition as to why we accept certain conclusions and not others. Finally, the example illustrates how to apply this approach to formalising the analysis of strong permission by Glavaničová [33].

We have also argued for many future research directions that may involve applying our method of building defeasible deontic logics. It is possible to handle various deontic challenges related to contrary-to-duty obligations, deontic detachment, and the formalism and legal interpretation of the right to privacy and the right to know if we extend the modal language properly in the ASPIC⁺-based defeasible logics. The essential step is to have ‘correct’ preference among formulas in the logic, or arguments in ASPIC⁺. What this ‘correctness’ is highly depends on what one wants to capture in the modelling. We are considering having a general approach to computing the construction of preference and then defeasible inference that may be based on, for instance, certain linguistic theories or legal theories.

Acknowledgments

We would like to thank the three referees for their valuable remarks and comments. Huimin Dong is supported by the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (20221187). Beishui Liao is supported by the Key Program of the National Social Science Foundation of China (20&ZD047). Leendert van der Torre also acknowledges financial support from the Fonds National de la Recherche Luxembourg (INTER/Mobility/19/13995684/DLAI/van der Torre).

This work was supported by the Fonds National de la Recherche Luxembourg through the project Deontic Logic for Epistemic Rights (OPEN O20/14776480).

References

- [1] Carlos E Alchourrón and Eugenio Bulygin. Permission and permissive norms. *Theorie der Normen*, pages 349–371, 1984.
- [2] Leila Amgoud and Philippe Besnard. Logical limits of abstract argumentation frameworks. *Journal of Applied Non-Classical Logics*, 23(3):229–267, 2013.
- [3] Albert J.J. Anglberger, Nobert Gratzl, and Olivier Roy. Obligation, free choice, and the logic of weakest permissions. *The Review of Symbolic Logic*, 8:807–827, December 2015.
- [4] Ofer Arieli, AnneMarie Borg, and Christian Straßer. Reasoning with maximal consistency by argumentative approaches. *Journal of Logic and Computation*, 28(7):1523–1563, 2018.
- [5] Nicholas Asher and Daniel Bonevac. Free choice permission is strong permission. *Synthese*, 145(3):303–323, 2005.
- [6] Guillaume Aucher, Guido Boella, and Leendert van der Torre. A dynamic logic for privacy compliance. *Artificial Intelligence and Law*, 19(2-3):187, 2011.
- [7] Chris Barker. Free choice permission as resource-sensitive reasoning. *Semantics and Pragmatics*, 3:10:1–38, 2010.
- [8] Diderik Batens. A strengthening of the rescher–manor consequence relations. *Logique et Analyse*, pages 289–313, 2003.
- [9] Diderik Batens, Joke Meheus, Dagmar Provijn, and Liza Verhoeven. Some adaptive logics for diagnosis. *Logic and Logical Philosophy*, 11:39–65, 2003.
- [10] Mathieu Beirlaen, Jesse Heyninck, and Christian Straßer. Structured argumentation with prioritized conditional obligations and permissions. *Journal of Logic and Computation*, 29(2):187–214, 2018.
- [11] Mathieu Beirlaen and Christian Straßer. Two adaptive logics of norm-propositions. *Journal of Applied Logic*, 11(2):147–168, 2013.
- [12] Mathieu Beirlaen, Christian Straßer, and Joke Meheus. An inconsistency-adaptive deontic logic for normative conflicts. *Journal of Philosophical Logic*, 42(2):285–315, 2013.
- [13] Salem Benferhat, Didier Dubois, and Henri Prade. A local approach to reasoning under inconsistency in stratified knowledge bases. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, volume 946, pages 36–43. Springer, 1995.
- [14] Patrick Blackburn, Maarten De Rijke, and Yde Venema. *Modal Logic*, volume 53. Cambridge University Press, 2002.
- [15] Jan Broersen. Deontic epistemic stit logic distinguishing modes of mens rea. *Journal of Applied Logic*, 9(2):137–152, 2011.

- [16] Martin Caminada. Rationality postulates: applying argumentation theory for non-monotonic reasoning. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of formal argumentation*. College Publication, 2018.
- [17] Claudette Cayrol. On the relation between argumentation and non-monotonic coherence-based entailment. In *International Joint Conference on Artificial Intelligence*, volume 95, pages 1443–1448, 1995.
- [18] R.M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, 24:33–36, 1963.
- [19] Amit Chopra, Leendert van der Torre, Harko Verhagen, and Serena Villata, editors. *Handbook of normative multiagent systems*. College Publications, 2018.
- [20] C. Condoravdi and S. Lauer. Anankastic conditionals are just conditionals. *Semantics and Pragmatics*, 9(8):1–69, November 2016.
- [21] Newton CA Da Costa and Walter A Carnielli. On paraconsistent deontic logic. *Philosophia*, 16(3-4):293–305, 1986.
- [22] Célia da Costa Pereira, Beishui Liao, Alessandra Malerba, Antonino Rotolo, Andrea G. B. Tettamanzi, Leendert W. N. van der Torre, and Serena Villata. Handling norms in multi-agent systems by means of formal argumentation. *FLAP*, 4(9):3039–3073, 2017.
- [23] Célia da Costa Pereira, Beishui Liao, Alessandra Malerba, Antonino Rotolo, Andrea GB Tettamanzi, Leendert van der Torre, and Serena Villata. Handling norms in multi-agent systems by means of formal argumentation. In Amit Chopra, Leendert van der Torre, Harko Verhagen, and Serena Villata, editors, *Handbook of normative multiagent systems*. College Publications, 2018.
- [24] Huimin Dong, Norbert Gratzl, and Olivier Roy. Open reading and free choice permission: A perspective in substructural logics. In *Dynamics, Uncertainty and Reasoning*, pages 81–115. Springer, 2019.
- [25] Huimin Dong, Xu Li, and Yi N. Wáng. Weighted modal logic in epistemic and deontic contexts. In Sujata Ghosh and Thomas Icard, editors, *Logic, Rationality, and Interaction*, pages 73–87. Springer International Publishing, 2021.
- [26] Huimin Dong, Beishui Liao, Réka Markovich, and Leendert W. N. van der Torre. From classical to non-monotonic deontic logic using aspic^+ . In *Logic, Rationality, and Interaction - 7th International Workshop, LORI 2019, Chongqing, China, October 18-21, 2019, Proceedings*, pages 71–85, 2019.
- [27] Huimin Dong, Réka Markovich, and Leendert van der Torre. Developing ai logic for social reasoning. *Journal of Zhejiang University*, 5(50):31–50, 2020.
- [28] Huimin Dong and Olivier Roy. Dynamic logic of power and immunity. In *International Workshop on Logic, Rationality and Interaction*, pages 123–136. Springer, 2017.
- [29] Huimin Dong and Olivier Roy. Dynamic logic of legal competences. *Journal of Logic, Language and Information*, 30(4):701–724, 2021.
- [30] Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors. *Handbook of deontic logic and normative systems: Volume 1*. College

- Publications, 2013.
- [31] Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors. *Handbook of deontic logic and normative systems: Volume 2*. College Publications, 2021.
 - [32] Dov M. Gabbay. Fibred semantics and the weaving of logics. part 1: Modal and intuitionistic logics. *The Journal of Symbolic Logic*, 61(4):1057–1120, 1996.
 - [33] Daniela Glavaničová. The free choice principle as a default rule. *Organon F*, 25(4):495–516, 2018.
 - [34] Lou Goble. Deontic logic (adapted) for normative conflicts. *Logic Journal of the IGPL*, 22(2):206–235, 2014.
 - [35] Guido Governatori. Practical normative reasoning with defeasible deontic logic. In *Reasoning Web International Summer School*, pages 1–25. Springer, 2018.
 - [36] Guido Governatori, Francesco Olivieri, Antonino Rotolo, and Simone Scannapieco. Computing strong and weak permissions in defeasible logic. *Journal of Philosophical Logic*, 42(6):799–829, 2013.
 - [37] Guido Governatori and Antonino Rotolo. Logic of violations: A gntzen system for reasoning with contrary-to-duty obligations. *The Australasian Journal of Logic*, 4, 2006.
 - [38] Guido Governatori and Antonino Rotolo. Is free choice permission admissible in classical deontic logic? *arXiv preprint arXiv:1905.07696*, 2019.
 - [39] Guido Governatori, Antonino Rotolo, and Erica Calardo. Possible world semantics for defeasible deontic logic. In *International Conference on Deontic Logic in Computer Science*, pages 46–60. Springer, 2012.
 - [40] Guido Governatori, Antonino Rotolo, and Giovanni Sartor. Deontics, logic, and the law. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*, volume 2. College Publication, 2020.
 - [41] Jörg Hansen. Conflicting imperatives and dyadic deontic logic. *Journal of Applied Logic*, 3(3-4):484–511, 2005.
 - [42] Jörg Hansen. Deontic logics for prioritized imperatives. *Artificial Intelligence and Law*, 14(1-2):1–34, 2006.
 - [43] Sven Ove Hansson. Preference-based deontic logic (PDL). *Journal of Philosophical Logic*, 19(1):75–93, 1990.
 - [44] Sven Ove Hansson. The varieties of permissions. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*. College Publication, 2013.
 - [45] Risto Hilpinen. Disjunctive permissions and conditionals with disjunctive antecedents. *Acta Philosophica Fennica*, 35:175–194, 1982.
 - [46] Wesley Newcomb Hohfeld. Fundamental legal conceptions applied in judicial reasoning. In Walter Wheeler Cook, editor, *Fundamental Legal Conceptions Applied in Judicial Reasoning and Other Legal Essays*, pages 23–64. New Haven : Yale University Press, 1923.

- [47] J. F. Horty. Nonmonotonic foundations for deontic logic. In D. Nute, editor, *Defeasible Deontic Logic*, pages 17–44. Kluwer, Dordrecht, 1997.
- [48] John F. Horty. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, 9(1-2):69–91, 1993.
- [49] John F Horty. Moral dilemmas and nonmonotonic logic. *Journal of philosophical logic*, 23(1):35–65, 1994.
- [50] John F Horty. *Agency and deontic logic*. Oxford University Press, 2001.
- [51] Hans Kamp. Free choice permission. In *Proceedings of the Aristotelian Society*, volume 74, pages 57–74. JSTOR, 1973.
- [52] Stig Kanger. Law and logic. *Theoria*, 38(3):105–132, 1972.
- [53] Stig Kanger and Helle Kanger. Rights and parliamentarism. *Theoria*, 32(2):85–115, 1966.
- [54] Sarit Kraus, Daniel J. Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44(1-2):167–207, 1990.
- [55] Ho-Pun Lam, Guido Governatori, and Régis Riveret. On aspic+ and defeasible logic. In *COMMA*, pages 359–370, 2016.
- [56] David Lewis. A problem about permission. In *Essays in honour of Jaakko Hintikka*, pages 163–175. Springer, 1979.
- [57] Beishui Liao, Nir Oren, Leender van der Torre, and Serena Villata. Prioritized norms in formal argumentation. *J. Log. Comput.*, 29(2):215–240, 2019.
- [58] Lars Lindahl. *Position and Change: A Study in Law and Logic*. Springer Science & Business Media, 1977.
- [59] D. Makinson and L. van der Torre. Constraints for input/output logics. *Journal of Philosophical Logic*, 30:155–185, 2001.
- [60] David Makinson. Stenius’ approach to disjunctive permission. *Theoria*, 50(2-3):138–147, 1984.
- [61] David Makinson. General patterns in nonmonotonic reasoning. In *Handbook of logic in artificial intelligence and logic programming (vol. 3)*, pages 35–110. Oxford University Press, 1994.
- [62] David Makinson. *Bridges from classical to nonmonotonic logic*. King’s College, 2005.
- [63] David Makinson and Leendert van der Torre. Permission from an input/output perspective. *Journal of philosophical logic*, 32(4):391–416, 2003.
- [64] Réka Markovich. Understanding Hohfeld and Formalizing Legal Rights: the Hohfeldian Conceptions and Their Conditional Consequences. *Studia Logica*, 108:129–158, 2020.
- [65] Sanjay Modgil and Henry Prakken. Abstract rule-based argumentation. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of formal argumentation*. College Publication, 2018.
- [66] D. Nute, editor. *Defeasible deontic logic*. Springer Netherlands, 1997.
- [67] Xavier Parent. Moral particularism in the light of deontic logic. *Artificial Intelligence and Law*, 19(2-3):75, 2011.

- [68] Xavier Parent and Leendert van der Torre. *Introduction to deontic logic and normative systems*. College Publications, 2018.
- [69] Xavier Parent and Leendert W. N. van der Torre. Detachment in normative systems: Examples, inference patterns, properties. *FLAP*, 4(9):2995–3038, 2017.
- [70] Gabriella Pigozzi and Leendert van der Torre. Arguing about constitutive and regulative norms. *Journal of Applied Non-Classical Logics*, 28(2-3):189–217, 2018.
- [71] Gabriella Pigozzi and Leendert W. N. van der Torre. Multiagent deontic logic and its challenges from a normative systems perspective. *FLAP*, 4(9):2929–2993, 2017.
- [72] Henry Prakken. Two approaches to the formalisation of defeasible deontic reasoning. *Studia Logica*, 57(1):73–90, 1996.
- [73] Henry Prakken and Giovanni Sartor. Formalising arguments about norms. In *Legal Knowledge and Information Systems (JURIX 2013)*, pages 121–130. IOS Press, 2013.
- [74] Henry Prakken and Marek Sergot. Contrary-to-duty obligations. *Studia Logica*, 57(1):91–115, 1996.
- [75] Joseph Raz. Permissions and supererogation. *American Philosophical Quarterly*, 12(2):161–168, 1975.
- [76] Régis Riveret, Antonino Rotolo, and Giovanni Sartor. A deontic argumentation framework towards doctrine reification. *Journal of Applied Logics*, 6(5):903–940, 2019.
- [77] David Ross. *The right and the good*. Oxford University Press, 1930.
- [78] John R Searle. *The construction of social reality*. Penguin, London, 1996.
- [79] Marek Sergot. Normative positions. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*. College Publication, 2013.
- [80] Christian Straßer. A deontic logic framework allowing for factual detachment. In *Adaptive Logics for Defeasible Reasoning*, pages 297–333. Springer, 2014.
- [81] Christian Straßer and Ofer Arieli. Normative reasoning by sequent-based argumentation. *Journal of Logic and Computation*, 29(3):387–415, 2019.
- [82] J. van Benthem, D. Grossi, and F. Liu. Priority structures in deontic logic. *Theoria*, 80(2):116–152, 2014.
- [83] Johan van Benthem. Minimal deontic logics. *Bulletin of the Section of Logic*, 8(1):36–42, 1979.
- [84] L. van der Torre. *Reasoning about obligations: defeasibility in preference-based deontic logic*. PhD thesis, Erasmus University, 1997.
- [85] L. van der Torre and Y. Tan. An update semantics for prima facie obligations. In *Proceedings of The 17th European Conference on Artificial Intelligence*, pages 38–42, 1998.
- [86] L. van der Torre and Y. Tan. Rights, duties and commitments between agents. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI 99, Stockholm, Sweden, July 31 - August 6, 1999. 2 Volumes, 1450 pages*, pages 1239–1246, 1999.

- [87] L. van der Torre and Y. Tan. An update semantics for defeasible obligations. In *UAI '99: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden, July 30 - August 1, 1999*, pages 631–638, 1999.
- [88] L. van der Torre and Y.-H. Tan. Two-phase deontic logic. *Logique et analyse*, 43(171-172):411–456, 2000.
- [89] B.C. van Fraassen. Values and the heart command. *Journal of Philosophy*, 70:5–19, 1973.
- [90] Georg Henrik von Wright. Deontic logic. *Mind*, 1951.
- [91] Georg Henrik von Wright. *Norm and Action - A Logical Enquiry*. Routledge, 1963.
- [92] Georg Henrik Von Wright. *An essay in deontic logic and the general theory of action*. North-Holland Publishing Company, 1968.
- [93] Anthony P. Young, Sanjay Modgil, and Odinaldo Rodrigues. Prioritised default logic as rational argumentation. In *Proceedings of AAMAS 2016*, pages 626–634, 2016.

Appendix: Proofs

Proposition 12.1. Consider the deontic language \mathcal{L} and a pair of two monotonic logics $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$. Let AF , corresponding to $\langle AT, \leq^\tau \rangle$, be an abstract argumentation framework $(\mathcal{A}, \mathcal{D})$ such that AT is based on $(\mathbf{S}^-; \mathbf{S}^+)$, K is a knowledge base, and $\tau \in \{p, r\}$. Given a set $\Gamma \subseteq \mathcal{L}$ of formulas, we define:

- a stable set generated by Γ as $\{D \in \mathcal{A} \mid F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma)\}$;
- a proper set generated by Γ as $\bigcup_{i \in \omega} E_i$, such that

$$E_0 = \{D \in \mathcal{A} \mid F(D) \subseteq Cn_{\mathbf{S}^-}(\Gamma)\}$$

$$E_{n+1} = \begin{cases} E_n \cup \{D \in \mathcal{A}\}, & \text{if } F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma) \text{ and} \\ & F(D) \cup F(E_n) \text{ is } \mathbf{S}^- \text{-consistent;} \\ E_n, & \text{otherwise.} \end{cases}$$

1. When $\tau = p$, then E is a stable set generated by a $\Gamma \in MC_{\mathbf{S}^+}(K)$ iff E is a stable extension regarding K .
2. When $\tau = r$, E is a proper set generated by a $\Gamma \in MC_{\mathbf{S}^-}(K)$ iff E is a stable extension regarding K .

Proof. 1. For the case of $\tau = p$.

The left-to-right direction. Let E be the stable set generated by a $\Gamma \in MC_{\mathbf{S}^+}(K)$.

- E is conflict-free. Otherwise there are $A, B \in E$ such that A defeats B . Suppose A rebuts B by the conclusion $\neg\varphi$ of a top rule of a subargument of B , such that $\text{Conc}(A) = +\varphi$. Then, from $F(A), F(B) \subseteq Cn_{\mathbf{S}^+}(\Gamma)$, we know that $-\varphi, +\varphi \in Cn_{\mathbf{S}^+}(\Gamma)$. This implies that Γ is not \mathbf{S}^+ -consistent, which contradicts $\Gamma \in MC_{\mathbf{S}^+}(K)$. When A undermines B , the result is the same.
- Given $B \notin E$, we need to find an $A \in E$ defeating B . We know that $F(B) \not\subseteq Cn_{\mathbf{S}^+}(\Gamma)$. Then, there is a $\varphi \in F(B)$ which is not derived from Γ in the system \mathbf{S}^+ . There are two cases to be considered.
 - φ is \mathbf{S}^+ -consistent with $Cn_{\mathbf{S}^+}(\Gamma)$. But then it contradicts the maximality of Γ .
 - φ is not \mathbf{S}^+ -consistent with $Cn_{\mathbf{S}^+}(\Gamma)$. Then, there are $\varphi_1, \dots, \varphi_n \in \Gamma$ such that $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^+} \neg\varphi$. So, there is an argument $A \in E$ with the top rule $\varphi_1, \dots, \varphi_n \Rightarrow \neg\varphi$. Because $\tau = p$, the premise-based ordering \leq^p ensures that $A \not\prec \varphi$, and then A undermines B .

Then, E is a stable extension regarding K .

The right-to-left direction. Let E be a stable extension regarding K . Let $\Gamma = E \cap K$.

- We will show that $\Gamma \in MC_{\mathbf{S}^+}(K)$.
 - Γ is \mathbf{S}^+ -consistent. Otherwise, there are $\varphi_1, \dots, \varphi_n, \varphi \in \Gamma$ such that $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^+} \neg\varphi$. There is an argument A with the top rule $\varphi_1, \dots, \varphi_n \Rightarrow \neg\varphi$. If $A \notin E$ and from that E there is a stable extension, assume that there is a $B \in E$ defeating A by the conclusion $\neg\varphi_n$. But both B and φ_n are in E , which contradicts that E is conflict-free.
 - Γ is maximal. Otherwise, there is a $\varphi \in K/\Gamma$ such that φ is \mathbf{S}^+ -consistent with Γ . But then $\varphi \notin E$. There is an argument $A \in E$ undermining φ . Suppose the top rule of A is $\varphi_1, \dots, \varphi_n \Rightarrow \neg\varphi$, where $\varphi_1, \dots, \varphi_n \in \Gamma$. It concludes that φ is not \mathbf{S}^+ -consistent with Γ .
- Let $D \in E$. We will show $F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma)$. If not, then there must be some $\varphi \in F(D)$ such that $\varphi \notin Cn_{\mathbf{S}^+}(\Gamma)$. However, this leads to a contradiction when we bring together the maximality of $\Gamma \in MC_{\mathbf{S}^+}(K)$ and the way it constructed $\Gamma = E \cap K$.
- Let $F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma)$. We will show that $D \in E$. Otherwise, there is an argument $A \in E$ defeating D by A 's top rule: $\varphi_1, \dots, \varphi_n \Rightarrow \neg\varphi$,

where $\varphi_1, \dots, \varphi_n \in Cn_{\mathbf{S}^+}(\Gamma)$ and $\varphi \in F(D)$. Then, $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^+} \neg\varphi$. However, by this assumption, we have $\neg\varphi, \varphi \in Cn_{\mathbf{S}^+}(\Gamma)$. And that contradicts the consistency of Γ .

2. For the case of $\tau = r$.

The left-to-right direction. Let E be a proper set generated by $\Gamma \in MC_{\mathbf{S}^-}(K)$.

- E is conflict-free. Otherwise, there are $A, B \in E$ such that A defeats B . There are four cases to be considered:
 - when both $A, B \in E_0$. Let $A = +\varphi$ and $B = -\varphi$. Then A undermines B . This implies $\Gamma \vdash_{\mathbf{S}^-} \perp$, which contradicts the consistency of Γ .
 - when $A = \varphi \in E_0$ and $B \in E_n$ ($n > 0$). We assume A rebuts B by the conclusion $\neg\varphi$ of the top rule of a subargument D of B . That then conflicts with the requirement that $F(B) \cup Cn_{\mathbf{S}^-}(\Gamma)$ being \mathbf{S}^- -consistent.
 - when $A \in E_n$ and $B \in E_0$ ($n > 0$). It is not possible for A to defeat B .
 - when $A \in E_n$ and $B \in E_m$ ($n, m > 0$). Suppose $n < m$. We then know that $A \in E_n \subseteq E_{m-1}$. We assume that A rebuts B by the conclusion $\neg\varphi$ of the top rule of a subargument D of B . That contradicts the requirement that $F(B) \cup F(E_{m-1})$ should be \mathbf{S}^- -consistent.
- For each $B \notin E$, we need to find a $A \in E$ defeating B . Consider:
 - when $B \in K/E$. Suppose $B = \varphi \in K$ and $B \notin \Gamma$. From $\Gamma \in MC_{\mathbf{S}^-}(K)$, it implies that $\Gamma \cup \{\varphi\}$ is not \mathbf{S}^- -consistent. There are $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^-} \neg\varphi$ where $\varphi_1, \dots, \varphi_n \in \Gamma$. Let $\varphi_1, \dots, \varphi_n \mapsto \neg\varphi$ be the top rule of an argument A . Then A undermines B and $A \in E$.
 - when $B \notin K$ and $B \notin E$.
 - * Suppose $\varphi \in F(B)/Cn_{\mathbf{S}^-}(\Gamma)$ and φ is the conclusion of a top rule in R_s . If φ is \mathbf{S}^- -consistent with $Cn_{\mathbf{S}^-}(\Gamma)$, that conflicts with the maximality of $\Gamma \in MC_{\mathbf{S}^-}(K)$. Then, there are $\varphi_1, \dots, \varphi_n \in \Gamma$ such that $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^-} \neg\varphi$. Let $\varphi_1, \dots, \varphi_n \mapsto \neg\varphi$ be the top rule of an argument $A \in E$. Then A undermines B .
 - * Suppose $\varphi \in F(B)/Cn_{\mathbf{S}^-}(\Gamma)$ such that φ is the conclusion of a top rule in R_d . Suppose φ is \mathbf{S}^- -consistent with $Cn_{\mathbf{S}^-}(\Gamma)$, which indicates that it is not possible to derive φ from Γ in system

\mathbf{S}^- . Then we can assume that $\varphi \in Cn_{\mathbf{S}^+}(\Gamma)$, and that φ is the only element in B to make $B \notin E$, i.e. φ is not \mathbf{S}^- -consistent with $Cn_{\mathbf{S}^-}(\Gamma)$. Then, there are $\varphi_1, \dots, \varphi_n \in Cn_{\mathbf{S}^-}(\Gamma)$ such that $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^-} \neg\varphi$. According to the definition of E_0 , there is an argument $A \in E_0$ with the top rule of $\varphi_1, \dots, \varphi_n \mapsto \neg\varphi$ rebutting B .

Then E is a stable extension regarding K .

The right-to-left direction. Let E be a stable extension regarding K . Let $\Gamma = Cn_{\mathbf{S}^-}(E \cap K)$.

- First of all, $\Gamma \in MC_{\mathbf{S}^-}(K)$.
 - Γ is \mathbf{S}^- -consistent. Otherwise, there are $\varphi_1, \dots, \varphi_n, \varphi \in \Gamma$ such that $\{\varphi_1, \dots, \varphi_n\} \vdash_{\mathbf{S}^-} \neg\varphi$. We have an argument $A = \varphi_1, \dots, \varphi_n \mapsto \neg\varphi$. If $A \in E$, then there is an argument $B = \varphi \in E$ such that A undermines B , which conflicts with E being conflict-free. If $A \notin E$, then because E is a stable extension, there is a $C = \psi_1, \dots, \psi_m \mapsto \neg\varphi_n \in E$ which undermines A for knowledge $\varphi_n \in Prem(A)$ where $\psi_1, \dots, \psi_m \in E$ (otherwise these premises would be defeated by some arguments contained in E , which contradicts that E is conflict-free). However, φ_n is already contained in E . This makes E not conflict-free. Given these two results, we know that Γ is \mathbf{S}^- -consistent.
 - Suppose Γ is not maximal. Let $\varphi \in K/\Gamma$. Then $\varphi \in K/E$. Because E is a stable extension, from $\varphi \notin E$ there is an $A \in E$ undermining φ . Assume that there is a top rule $\varphi_1, \dots, \varphi_n \mapsto \neg\varphi$ of a subargument of A where $\varphi_1, \dots, \varphi_n \in \Gamma$. Then $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^-} \neg\varphi$. That implies that $\Gamma \cup \{\varphi\}$ is not \mathbf{S}^- -consistent.
- Given any $D \in E$, we show that either $D \in E_0$ or $D \in E_{n+1}$ ($n \geq 0$). Suppose $D \in E/E_0$. We show that $D \in E_{n+1}$ for some $n \geq 0$. Otherwise, given any E_n ($n \geq 0$), assume that $F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma)$ but there is a $\varphi \in F(D)$ such that φ is not \mathbf{S}^- -consistent with $F(E_n)$. There are $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^-} \neg\varphi$ where $\varphi_1, \dots, \varphi_n \in F(E_n)$. So, there is an argument $A \in E_n$ with the top rule $\varphi_1, \dots, \varphi_n \mapsto \neg\varphi$ in R_s , which defeats D . This conflicts with E being conflict-free.
- Given $D \in E_0$ or $D \in E_{n+1}$ ($n \geq 0$), we show $D \in E$. Consider:
 - when $D \in E_0$. Then by the way a stable set is defined, D is contained in E .

- when $D \in E_{n+1}$ ($n \geq 0$) with $D \notin E_n$. Suppose $E_n \subseteq E$ and suppose $D \notin E$. Then, from that E is a stable extension, and there is an argument $A \in E$ defeating D . Let the top rule of A be $\varphi_1, \dots, \varphi_n \Rightarrow \neg\varphi$ where $\varphi_1, \dots, \varphi_n \in Cn_{\mathbf{S}^+}(\Gamma)$ and $\varphi \in F(D)$. Then $\varphi \notin Cn_{\mathbf{S}^+}(\Gamma)$. Then $D \notin E_{n+1}$.

Then E is a proper set generated by $\Gamma \in MC_{\mathbf{S}^-}(K)$. □

Proposition 12.2. Let $\Gamma \subseteq \mathcal{L}$, $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics and let K be a knowledge base of AT . We define

- an R-set generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$ as $\bigcup_{n \in \mathbb{N}} R_n$, such that:

$$R_0 = Cn_{\mathbf{S}^-}(\Gamma)$$

$$R_{n+1} = \begin{cases} R_n \cup \{\varphi\}, & \text{if } \varphi \in Cn_{\mathbf{S}^+}(\Gamma) \text{ and} \\ & \{\varphi\} \cup R_n \text{ is } \mathbf{S}^- \text{-consistent;} \\ R_n, & \text{otherwise;} \end{cases}$$

where $\Gamma \in MC_{\mathbf{S}^-}(K)$.

The R-collection $R_{\mathbf{S}^-; \mathbf{S}^+}(K)$ generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$ is the set of all R-sets generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$. Then:

1. $\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{p\forall}(K) = \bigcap_{\Gamma \in MC_{\mathbf{S}^+}(K)} Cn_{\mathbf{S}^+}(\Gamma)$;
2. $\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{r\forall}(K) = \bigcap R_{\mathbf{S}^-; \mathbf{S}^+}(K)$.

Proof. This proposition can be a direct result from Proposition 7.1.

1. When $\varphi \in \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{p\forall}(K)$, then there is an argument A with $Conc(A) = \varphi$ such that A is contained in every stable extension regarding K . By Proposition 7.1.1, A is contained in every stable set generated by $\Gamma \in MC_{\mathbf{S}^+}(K)$, and then $\varphi \in Cn_{\mathbf{S}^+}(\Gamma)$ for every $\Gamma \in MC_{\mathbf{S}^+}(K)$. This result leads to $\varphi \in \bigcap_{\Gamma \in MC_{\mathbf{S}^+}(K)} Cn_{\mathbf{S}^+}(\Gamma)$. The other direction is similar.
2. When $\varphi \in \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{r\forall}(K)$, as in the previous case, by applying Proposition 7.1.2, we reach the same result. The other direction is similar by taking the definition into consideration. □

Proposition 12.3. Consider the deontic language \mathcal{L} and a pair of two monotonic logics $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$. Let AF , corresponding to $\langle AT, \leq^\tau \rangle$, be an abstract argumentation framework $(\mathcal{A}, \mathcal{D})$ such that AT is based on $(\mathbf{S}^-; \mathbf{S}^+)$, $K = K_s \cup K_d$ is a knowledge base, and $\tau \in \{f, o^s, a^s, o, a, d, fr\}$. We construct a τ -premise set generated by K as $\bigcup_{n \in \mathbb{N}} E_n$ such that:

$$E_0 = \{D \in \mathcal{A} \mid F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma_1)\} \text{ for some } \Gamma_1 \in MC_{\mathbf{S}^+}(K^\tau)$$

$$E_{n+1} = \begin{cases} E_n \cup \{D \in \mathcal{A}\}, & \text{if } \exists \Gamma_2 \in MC_{\mathbf{S}^+}(K - K^\tau) \text{ such that} \\ & (i) F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma_2) \text{ and} \\ & (ii) F(D) \cup F(E_n) \text{ is } \mathbf{S}^+\text{-consistent;} \\ E_n, & \text{otherwise.} \end{cases}$$

Then:

- E is a τ -premise set generated by K iff E is a stable extension regarding K .

Proof. The proof in this proposition is similar to the proof strategy in Proposition 7.1.

The left-to-right direction. Let $E = \bigcup_{n \in \mathbb{N}} E_n$ be a τ -premise set generated by K .

- E is conflict-free. Otherwise, there are $A, B \in E$ such that A defeats B . Consider:
 1. when $A, B \in E_0$. Then $+\varphi, -\varphi \in Cn_{\mathbf{S}^+}(\Gamma_1)$ where $\Gamma_1 \in MC_{\mathbf{S}^+}(K^\tau)$. But then that leads to a contradiction of Γ_1 , which is \mathbf{S}^+ -consistent.
 2. when $A, B \notin E_0$. We assume that $A \in E_m$ and $B \in E_n$ with $m < n$. By the construction of E , we can simply suppose that $B \in E_{m+1}$. Suppose A undermines B by having $+\varphi = Conc(A) \in F(E_m)$ and $-\varphi = Prem(B')$. Then, $F(B) \cup F(E_m)$ is not \mathbf{S}^+ -consistent, which contradicts the construction of E .
- Given $B \notin E$, we need to find a $A \in E$ such that it defeats B . Consider:
 1. when $B \in K/E$. Suppose $B = \varphi \in K$ and $B \notin \Gamma$ where $\Gamma \in MC_{\mathbf{S}^+}(K^\tau)$. That simply implies that $\Gamma \cup \{\varphi\}$ is not \mathbf{S}^+ -consistent. Then, there is $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^+} \neg\varphi$ with $\varphi_1, \dots, \varphi_n \in \Gamma$. Let $\varphi_1, \dots, \varphi_n \mapsto \neg\varphi$ be the top rule of an argument A . Because all the premises of A and B come from K^τ , then A undermines B .
 2. when $B \notin K$ and $B \notin E$.

- (a) Suppose $\varphi \in F(B)/Cn_{\mathbf{S}^+}(\Gamma_1)$ where $\Gamma_1 \in MC_{\mathbf{S}^+}(K^\tau)$ such that φ is the conclusion of a top rule in R_s . Because of the maximality of Γ_1 , there are $\varphi_1, \dots, \varphi_n \in \Gamma_1$ such that $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^+} \neg\varphi$. Let $\varphi_1, \dots, \varphi_n \mapsto \neg\varphi$ be the top rule of an argument $A \in E_1$. Because all the premises of A come from K^τ , then A undermines B .
- (b) Suppose $\varphi \in F(B)/Cn_{\mathbf{S}^+}(\Gamma_1)$ where $\Gamma_1 \in MC_{\mathbf{S}^+}(K^\tau)$ such that φ is the conclusion of a top rule in R_d . We can still find such an $A \in E_1$, as above, that rebuts B .
- (c) Suppose $\varphi \in F(B)/Cn_{\mathbf{S}^+}(\Gamma_2)$ where $\Gamma_2 \in MC_{\mathbf{S}^+}(K - K^\tau)$ such that φ is the conclusion of a top rule (either in R_s or in R_d). The argument of proof is similar to that for the previous two cases.

Then E is a stable extension.

The right-to-left direction. Let E be a stable extension regarding K . Let $\Gamma = Cn_{\mathbf{S}^+}(E \cap K^\tau)$. Then:

- we will show that $\Gamma \in MC_{\mathbf{S}^+}(K^\tau)$.
 1. Γ is \mathbf{S}^+ -consistent. Otherwise, there are $\varphi_1, \dots, \varphi_n, \varphi \in \Gamma$ such that $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^+} \neg\varphi$. Then, we have an argument $A = \varphi_1, \dots, \varphi_n \Rightarrow \varphi$. If $A \in E$, then there is an argument $B = \varphi \in E$ such that A undermines B . This is because all the premises of A are in K^τ . But then, that conflicts with E being conflict-free. If $A \notin E$, then because E is a stable extension, there is a $C = \psi_1, \dots, \psi_m \mapsto \neg\varphi_n \in E$ which undermines A for knowledge $\varphi_n \in Prem(A)$ where $\psi_1, \dots, \psi_m \in E$. Then $Prem(C) \subseteq K^\tau$, otherwise C is not preferable enough to defeat A . However, φ_n is already contained in E . That leads to E not being conflict-free. In sum, Γ is \mathbf{S}^+ -consistent.
 2. Γ is maximal. Otherwise, let $\varphi \in K^\tau/\Gamma$ such that $\Gamma \cup \{\varphi\}$ is \mathbf{S}^+ -consistent. Then $\varphi \notin E_0$. So, $\varphi \in Cn_{\mathbf{S}^+}(\Gamma) = Cn_{\mathbf{S}^+}(K^\tau) = \Gamma$, which leads to a contradiction.
- given any $D \in E$, we will show that there is an $n \in \mathbb{N}$ such that $D \in E_n$. We prove this by induction on the structure of D . Consider:
 - when $D \in K$. Since $D \in E$ and $\Gamma = Cn_{\mathbf{S}^+}(E \cap K^\tau)$, from $D \in E \cap K$ it is implied that $D \in Cn_{\mathbf{S}^+}(\Gamma)$. So $D \in E_0$.
 - when $D = \mapsto \in R_s^0$ or $D = \mapsto \in R_d^0$. It is clear that $F(D) \subseteq Cn_{\mathbf{S}^+}(\Gamma)$.
 - when we have an inductive hypothesis. All the subarguments of D are contained in some E_n where $n \in \mathbb{N}$.

- when $D = D_1, \dots, D_n \mapsto \varphi$. Then $(\textcircled{a}) F(D_1), \dots, F(D_n) \vdash_{\mathbf{S}^-} \varphi$. Notice that $F(D_i)$ is \mathbf{S}^+ -consistent for each $i \in [1, n]$ by inductive hypothesis. If for all E_n it is the case that $D \notin E_n$, then $F(D_1) \cup \dots \cup F(D_n) \cup \{\varphi\}$ is not \mathbf{S}^+ -consistent. But then that conflicts with (\textcircled{a}) .
 - when $D = D_1, \dots, D_n \Rightarrow \varphi$. The proof is similar to that of the previous case.
- given a $D \notin E$, we will show that there is no E_n such that $D \in E_n$ is constructed from Γ . Since E is a stable extension, there is an argument $A \in E$ such that A defeats D . Then, there are $\varphi_1, \dots, \varphi_n \vdash_{\mathbf{S}^+} \varphi$ such that $\varphi_1, \dots, \varphi_n, \varphi \in F(A)$ and $\neg\varphi \in F(D)$. Then, $A \in E_n$ for some $n \in \mathbb{N}$ by the case proven in the previous step. Now $\neg\varphi \notin E_n$ for any $n \in \mathbb{N}$, otherwise the result will be contrary to Γ being \mathbf{S}^+ -consistent. It is then concluded that $D \notin E_n$ for any $n \in \mathbb{N}$.

□

Proposition 12.4. Consider the deontic language \mathcal{L} and a pair of two monotonic logics $(\mathbf{S}^-; \mathbf{S}^+)$. Let AF_i , corresponding to $\langle AT, \leq^i \rangle$, be an abstract argumentation framework $(\mathcal{A}, \mathcal{D}_i)$ such that AT is based on $(\mathbf{S}^-; \mathbf{S}^+)$, K is a knowledge base, and $i \in \{1, 2\}$. Let $Stable(AF_i)$ be the set of all stable extensions w.r.t. AF_i .

1. If $K^{\leq 1} \subseteq K^{\leq 2}$, then $E \in Stable(AF_1)$ implies $\exists E' \in Stable(AF_2)$ s.t. $E' = E$.
2. If $K^{\leq 1} \subseteq K^{\leq 2}$, then $|Stable(AF_1)| \leq |Stable(AF_2)|$.

Proof. 1. Consider the case when $K^{\leq 1} \subseteq K^{\leq 2}$. Suppose $E \in Stable(AF_1)$. Let $E = E_1 \cup E_2$ according to Proposition 8.1 and Proposition 8.3. We need to show that E is a stable extension w.r.t. AF_2 .

First, E is conflict-free in AF_2 . Otherwise $\exists A, B \in E$ such that $(A, B) \in \mathcal{D}$ in AF_1 . If A rebuts B , then $Conc(A) = \neg\varphi$ for some $B' \in Sub(B)$ and $TopRule(B') \in R_d$, $Con(B') = \varphi$, and $A \not\prec B'$. So B is generated from $K - K^{\leq 2}$ and then from $K - K^{\leq 1}$. This indicates that $(A, B) \in \mathcal{D}$ in AF_2 . But then that contradicts E being conflict-free in AF_1 . If A undermines B , then $Conc(A) = \neg\varphi$ for knowledge $\varphi \in Prem(B)$ of B and $A \not\prec \varphi$. No matter where A is generated from, whether from $K^{\leq 1}$ or from $K - K^{\leq 1}$, it keeps the preference in $K^{\leq 2}$. So A undermines B in AF_2 . Again, this contradicts E being conflict-free in AF_1 . Thus, we conclude that E is conflict-free in AF_2 .

Now, we prove that $\forall B \notin E \exists A \in E$ such that $(A, B) \in \mathcal{D}$ in AF_2 . If that is not the case, then $\exists B \notin E$ such that $\forall A \in E$ and $(A, B) \notin \mathcal{D}$ in AF_2 (\textcircled{a}) .

Notice that from E there is a stable extension in AF_1 . We then have $\exists A' \in E$ such that $(A', B) \in \mathcal{D}$ in AF_1 . Since $K^{\leq^1} \subseteq K^{\leq^2}$, this implies that $(A', B) \in \mathcal{D}$ in AF_2 . That conflicts with $(@)$. So the assumption is false.

Now we conclude that E is a stable extension in AF_2 .

2. From the first item, we can easily see that this statement holds. Notice that we have

$$\leq^{o^s} \subseteq \leq^o \quad \text{and} \quad \leq^{a^s} \subseteq \leq^a \quad \text{and} \quad \leq^f \subseteq \leq^o \quad \text{and} \quad \leq^f \subseteq \leq^a .$$

And thus it implies:

- $|Stable(\langle AT, \leq^{o^s} \rangle)| \leq |Stable(\langle AT, \leq^o \rangle)|$;
- $|Stable(\langle AT, \leq^{a^s} \rangle)| \leq |Stable(\langle AT, \leq^a \rangle)|$;
- $|Stable(\langle AT, \leq^f \rangle)| \leq |Stable(\langle AT, \leq^o \rangle)|$;
- $|Stable(\langle AT, \leq^f \rangle)| \leq |Stable(\langle AT, \leq^a \rangle)|$.

□

Proposition 12.5. Let $\Gamma \subseteq \mathcal{L}$, $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics, let \leq^τ be a τ -ordering ($\tau \in \{p, r, f, o^s, a^s, o, a, d, fr\}$), and let K be a knowledge base of AT . We define

- a P-set generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$ as $\bigcup_{n \in \mathbb{N}} P_n$, such that

$$P_0 = Cn_{\mathbf{S}^+}(\Gamma)$$

$$P_{n+1} = \begin{cases} P_n \cup \{\varphi\}, & \text{if } \exists \Gamma' \in MC_{\mathbf{S}^+}(K - K^\tau) \text{ such that} \\ & (i) \varphi \in Cn_{\mathbf{S}^+}(\Gamma') \text{ and} \\ & (ii) \{\varphi\} \cup P_n \text{ is } \mathbf{S}^+\text{-consistent;} \\ P_n, & \text{otherwise;} \end{cases}$$

where $\Gamma \in MC_{\mathbf{S}^+}(K^\tau)$.

The P-collection $P_{\mathbf{S}^-, \mathbf{S}^+}(K)$ generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$ is the set of all P-sets generated by K in $(\mathbf{S}^-; \mathbf{S}^+)$. Then

- $C_{\mathbf{S}^-, \mathbf{S}^+}^{\tau \forall}(K) = \bigcap P_{\mathbf{S}^-, \mathbf{S}^+}(K)$.

Proof. Just like the proof in Proposition 7.2, by applying Proposition 8.1 and Proposition 8.3, the result can be reached. □

Proposition 12.6. Let $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics. Now we have $p \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau \forall} p$ but $\{p, \neg p\} \not\vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau \forall} p$, with $\tau \in \{p, r, f, o^s, a^s, o, a, d, fr\}$.

Proof. This proposition can be argued by applying Proposition 7.2, Proposition 8.1 and Proposition 8.3. \square

Proposition 12.7. Let $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ be a pair of two monotonic logics. We have the following relations regarding supra-classicality:

$$\vdash_{\mathbf{S}^-} \subseteq \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \subseteq \vdash_{\mathbf{S}^+}$$

where $\tau \in \{p, r, f, o^s, a^s, o, a, d, fr\}$.

Proof. Again, this proposition can be proved by applying Proposition 7.2, Proposition 8.1 and Proposition 8.3. \square

Proposition 12.8. Given $\tau \in \{p, r, f, o^s, a^s, o, a, d, fr\}$ as one of the preferences defined and $(\mathbf{S}^-; \mathbf{S}^+) \in \{(\mathbf{D}_{-2}; \mathbf{D}_{-1}), (\mathbf{D}_{-2}; \mathbf{D}), (\mathbf{D}_{-1}; \mathbf{D})\}$ as a pair of two monotonic logics, we will now check whether the defeasible deontic logics defined in this article satisfy the following standard properties regarding non-monotonicity (where we simplify $\|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}$ to \Vdash):

1. Reflexivity: $\Gamma \Vdash \varphi$ where $\varphi \in \Gamma$
2. Cut: If $\Gamma \cup \{\psi\} \Vdash \chi$ and $\Gamma \Vdash \psi$, then $\Gamma \Vdash \chi$
3. Cautious Monotony: if $\Gamma \Vdash \psi$ and $\Gamma \Vdash \chi$, then $\Gamma \cup \{\psi\} \Vdash \chi$
4. Left Logical Equivalence: if $Cn_{\mathbf{S}^+}(\Gamma) = Cn_{\mathbf{S}^+}(\Gamma')$ and $\Gamma \Vdash \chi$, then $\Gamma' \Vdash \chi$
5. Right Weakening: if $\vdash_{\mathbf{S}^+} \varphi \rightarrow \psi$ and $\Gamma \Vdash \varphi$, then $\Gamma \Vdash \psi$
6. OR: if $\Gamma \Vdash \varphi$ and $\Gamma' \Vdash \varphi$, then $\Gamma \cup \Gamma' \Vdash \varphi$
7. AND: if $\Gamma \Vdash \psi$ and $\Gamma \Vdash \chi$, then $\Gamma \Vdash \psi \wedge \chi$
8. Rational Monotony: If $\Gamma \Vdash \chi$ and $\Gamma \not\vdash \neg\psi$, then $\Gamma \cup \{\psi\} \Vdash \chi$

The results are shown in Table 8.

Proof. We first check whether the following properties hold for $\|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}$ by using Propositions 7.2 and 8.4. Consider:

1. Reflexivity (for the rule-based preference). First we know that for the rule-based preference, all the different knowledge are better than the other arguments. According to the construction shown in Proposition 7.2 and 8.4, we can see that all the different knowledge are contained in the consequences. Thus Reflexivity holds.

Properties	$\ \sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}$	$\ \sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}$
Reflexivity	\checkmark^*	No
Cut	\checkmark	\checkmark
Cautious Monotony	\checkmark	\checkmark
Left Logical Equivalence	\checkmark	\checkmark
Right Weakening	No	\checkmark
OR	No	No
AND	\checkmark	\checkmark
Rational Monotony	\checkmark	\checkmark

Table 8: This is a summary of various consequences we have based on different types of knowledge base. Notice that $\tau \in \{p, f, o^s, a^s, o, a, d, fr\}$. The symbol \checkmark^* indicates that this property is satisfied when the given knowledge base is consistent in \mathbf{S}^- .

2. Cut. Suppose $\Gamma \cup \{\psi\} \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \chi$ and $\Gamma \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \psi$. From the latter, we have $\psi \in \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma)$. This implies that $\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\{\Gamma \cup \{\psi\}\}) \subseteq \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma)$. By applying the first assumption, we conclude that $\chi \in \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma)$.
3. Cautious Monotony. Assume that $\Gamma \cup \{\psi\} \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \chi$ and $\Gamma \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \psi$. From the second assumption, we get $\psi \in \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma)$. This indicates that $\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma \cup \{\psi\}) \subseteq \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma)$. Applying this result to the second assumption, we then conclude that $\chi \in \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma)$.
4. Left Logical Equivalence. Assume that $Cn(\Gamma) = Cn(\Gamma')$ and $\Gamma \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \chi$. From the second assumption, we then have $\chi \in \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma)$. Because $Cn(\Gamma) = Cn(\Gamma')$, it is implied that $\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma) = \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma')$. This immediately indicates that $\chi \in \mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma')$.
5. Right Weakening (for the preferences based on premises). Suppose $\vdash_{\mathbf{S}^+} \varphi \rightarrow \psi$ and $\Gamma \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \varphi$. By $\Gamma \|\sim_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall} \varphi$ and Proposition 8.4, there are two cases of φ being contained in every stable extension. If φ is contained in the maximally consistent subset regarding K^τ , then ψ is also contained in K^τ because $\vdash_{\mathbf{S}^+} \varphi \rightarrow \psi$. By the construction of Proposition 8.4, we know that ψ is one of the best arguments, and thus cannot be defeated. It is also contained in $\mathcal{C}_{\mathbf{S}^-; \mathbf{S}^+}^{\tau \forall}(\Gamma)$. If φ is contained in the maximally consistent subset regarding $K - K^\tau$, then φ is always the conclusion of one of the second best arguments. Suppose such an argument is $A \in E' \in \text{Stable}(\Gamma)$ such that $\text{Conc}(A) = \varphi$ for any stable

extension E' (@). Notice that $Prem(A) \subseteq K - K^\tau$. Let $E \in Stable(\Gamma)$ be a stable extension. Because $\vdash_{\mathbf{S}^+} \varphi \rightarrow \psi$, we then have an argument $A' = A \Rightarrow \psi$ where $TopRule(A') = \varphi \Rightarrow \psi$. Because A is contained in a stable extension E , the only way to defeat A' is to rebut it by the conclusion ψ . This indicates that there is an argument $B \in E$ such that $Conc(B) = \neg\psi$ and $B \not\prec A'$. On the other hand, we then have an argument $B' = B \Rightarrow \neg\varphi$. Notice that $Prem(A) = Prem(A')$ and $Prem(B) = Prem(B')$. Thus $B' \not\prec A$ because of the preference on premises. So A is defeated by B' . This indicates that there is a stable extension containing B' and thus excludes A . But then that conflicts with (@). So A' is not defeated in E . We then conclude that $\Gamma \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall} \psi$.

6. AND. Assume that $\Gamma \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall} \psi$ and $\Gamma \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall} \chi$. So $\psi \in \mathcal{C}_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall}(\Gamma)$ and $\chi \in \mathcal{C}_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall}(\Gamma)$. We can have $\psi \wedge \chi \in \mathcal{C}_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall}(\Gamma)$.
7. Rational Monotony. Assume that $\Gamma \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall} \chi$ and $\Gamma \not\vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall} \neg\psi$. We want to show that $\Gamma \cup \{\psi\} \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall} \chi$. By $\Gamma \not\vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall} \neg\psi$, there is a stable extension $E \in Stable(\Gamma)$ such that for any argument $A \in E$, then $Conc(A) \neq \neg\psi$ (@).
 - (a) If ψ is the conclusion of one argument in a stable extension of the Γ -AT, this indicates that ψ is \mathbf{S}^+ -consistent with the conclusion χ of an argument in this stable extension. So adding ψ to the knowledge base Γ does not change its stable extensions. Then we have the desired result.
 - (b) If ψ is the conclusion of one argument that is excluded in every stable extension of the Γ -AT, then by the assumption (@), all the conclusions of all the arguments in E are \mathbf{S}^+ -consistent with ψ . After adding ψ to the knowledge base, Γ does not change its stable extensions. Then we have the desired result again.
 - (c) If there is no argument from Γ has ψ as its conclusion, then adding ψ to the knowledge base Γ does not change its stable extensions. Then we have the desired result again.

According to the above cases, we conclude that $\Gamma \cup \{\psi\} \Vdash_{\mathbf{S}^-, \mathbf{S}^+}^{\tau\forall} \chi$.

□

BUSINESS PROCESS MODELLING IN HEALTHCARE AND COMPLIANCE MANAGEMENT: A LOGICAL FRAMEWORK

ILARIA ANGELA AMANTEA
University of Turin, Turin, Italy
amantea@di.unito.it

LIVIO ROBALDO
Legal Innovation Lab Wales
Hillary Rodham Clinton School of Law
Swansea University
livio.robaldoswansea.ac.uk

EMILIO SULIS
University of Turin, Turin, Italy
sulis@di.unito.it

GUIDO GOVERNATORI
CSIRO, Data61, QLD, Australia
guido.governatori@data61.csiro.au

GUIDO BOELLA
University of Turin, Turin, Italy
boella@di.unito.it

Abstract

This work describes a methodological approach to investigate Compliance Management in healthcare based on a BPM perspective, exploring an application in an innovative hospital service. Firstly, we present a business process analysis by modeling the process with the adoption of a standard language. Secondly, we encode a set of rules in LegalRuleML, an XML formalism designed

to be a standard for representing the semantic and logical content of legal documents. The rules represent some provisions of the General Data Protection Regulation (GDPR) that are involved in the health process analyzed. Moreover, in order to perform the regulatory compliance check automatically, we converted the set of rules into Defeasible Deontic Logic format (DDL), readable by the Regorous compliance checker developed at CSIRO. Overall, the paper shows a methodology to automate regulatory compliance checking of a real hospital process with actual regulations and norms. The codes in the LegalRuleML and DDL formats used in the work are available online¹.

1 Introduction

One of the main research topics in Business Process Management (BPM) concerns regulatory or Compliance Management (CM), i.e. the analysis of compliance to norms [26; 60]. The necessity of satisfying regulations or laws forces organizations in redesign their internal processes, in the context of change management [40]. The increasing pressure from regulatory authorities to organizations led to the development and application of Compliance Management Frameworks (CMFs). In this context, CM can be addressed at the operational level by focusing on business processes, intended as the set of activities accomplishing a specific organizational goal.

Business process analysis usually introduces performance objectives to be considered in addition to constraints imposed by external pressures (e.g., regulatory issues). The investigation of undesirable events and norm violations adopted traditional techniques, e.g. root cause analysis (commonly used in manufacturing processes to improve performance). More recently, CMFs explore the relationship between the formal representation of a process model and the relevant regulations. There are many different adoptable CM strategies consisting in approaches to check whether a business process complies with the actual regulation automatically [35]. The goal is to ensure that such approaches properly model business processes as well as norms. Moreover, in the past decades many CM approaches in the context of digitization to automatize business processes have been proposed [50]. We describe here a CM approach to support regulatory compliance for healthcare business processes based on a compliance-by-design methodology [35] and using a business process compliance checker called Regorous [28]. In particular, this paper explores the adoption of a two-step pipeline introducing a CMF applied to an innovative hospital service. In a first step, business process analysis can be performed by adopting standard modeling language to investigate healthcare processes at operational level. In a second step, a regulatory CM is proposed on the top of the model by applying a logic-based

¹See <https://github.com/liviorobaldo/BPMinHealthcare>

approach to automate checking whether the process complies with the new General Data Protection Regulation (GDPR).

The rest of this paper is structured as follows: background and related work, detail of the analysis of the case study on Business Process Management prospective and Regulatory Compliance prospective and finally, results and discussions on future work.

2 Framework and Related Work

2.1 Risk management and regulatory compliance

Risk is part of every business activity and therefore part of every business process [60; 39]. The occurrence of a risk may lead to loss of quality, increased costs, time delays, complaints, and legal problems [17] as well as, in healthcare, serious and permanent damages up to death. There are several types of risk, such as legal, procedural, economical, financial, etc. The Risk Management is the discipline that allows the management of these different kinds of risks thank to the application of some principles [51; 42; 36].

Regarding legal risks, it should be considered that the process has to be compliant to law, whereas norms and regulations are constantly evolving and new reorganizations must be implemented with the introduction of new procedures [40], i.e., for privacy control, AI technologies.

Compliance in healthcare considers the conformity of care processes with laws, regulations and standards related to patient safety, privacy of patient information and administrative practices [7; 44].

Ultimately, health compliance is about providing safe and high quality patient care. Healthcare organizations are also required to comply with strict standards, regulations and laws at regional and state level. Violations of these laws may result in legal action, heavy fines or loss of licenses.

It is possible to find several studies on compliance with laws, rules or regulations in the case of processes related to patient health [23; 48; 9; 5; 6].

The intensive use of ICT solutions to collect, share and digitize data of a health process, makes it necessary to prepare tools able to identify any possible risk scenario related to the use of computer systems and lack of awareness on the agents, as well as to facilitate the adoption of appropriate counter-measures. Previous research on IT in healthcare explored digitalization challenges for organization [Amantea et al., 2018]. These innovations may require the application of new regulations, such as the GDPR, without forgetting that the health sector is full of strict health regulations in constant evolution.

2.2 Business process compliance and logic

Regulatory compliance is the set of activities an enterprise undertakes to ensure that its core business does not violate relevant regulations, in the jurisdictions in which the business is situated, governing the (industry) sectors where the enterprise operates. The activities an organization undertakes to achieve its business objectives can be represented by the business processes of the company. On the other hand, a normative document (e.g., a code, a guide line, an act) can be understood as a set of clauses, and these clauses can be represented in an appropriate formal language.

2.2.1 Business process modeling

In order to analyse the use case hospital business processes, we exploit a Business Process Management (BPM) methodology. One of the central issues in BPM is change management [61; 1; 25]. Using a process-centric approach, in order to describe the diagram of the process, we will adopt the Business Process Model and Notations (BPMN) standard language [2]. Primarily, in the context of health-care studies, BPMN standard language acquires a peculiar consideration [43; 8; 56].

The business process analysis aims to define and engineer a model to be verified and validated with system experts. One of the main output is the creation of visual models of processes (i.e., process map or flowchart). These diagrams depict the sequence of activities and various crossroads (gateways), which lead to different routes depending on choices made. A business process model is a self-contained, temporal and logical order in which a set of activities are expected to be executed to achieve a business goal. Typically, a process model describes what needs to be done and when (control flow), who is going to do what (resources), and on what it is working on (data). In this context, a possible execution, called process trace or simply trace, is a sequence of tasks and events respecting the order given by the connectors.

2.2.2 The automation of compliance

Business process compliance is a relationship between the formal representation of a process model and the formal representation of the relevant regulations [33]. Any approach to automatically check whether a business process complies with the regulation governing has to ensure that it is able to properly model business processes as well as norms. In the past decades many approaches to automatize business process compliance have been proposed [41; 16] and legal informatics is experiencing growth in activity [20; 18; 15; 59; 45].

However, a challenging research topic is the possibility of modeling standards in a conceptually valid, detailed and exhaustive way that can be used in practice for companies and, at the same time, have the ability to be used generically for any type of standard also taking into account the regulatory environment as a whole [28].

This shifts the focus to the adoption formalisms. Temporal logic and Event Calculus have been used in several frameworks. However, it has been shown that when norms are formalized in Linear Temporal Logic the evaluation whether a process is compliant produces results that are not compatible with the intuitive and most natural legal interpretation [38; 29]. Furthermore, it was argued that, while such logics can properly model norms, such formalizations would be completely useless from a process compliance point of view insofar they would require an external oracle to identify the compliant executions of the process, and build the formalization from the traces corresponding to the traces deemed legal by the oracle. This means that, there is no need for the formalization to determine if the process is compliant or not, since this is done by the oracle [29; 31]. Some studies had focused on the application of Natural Language Processing (NLP) methods to design legal document management system to assist legal professionals in navigate legislation and retrieving the information they are interested in [21; 22]. An example is Eunomos [19; 18]. These types of systems classify, index, and discover inter-links between legal documents, retrieved through Web-crawling tools, by exploiting NLP tools, such as parsers and statistical algorithms, and semantic knowledge bases, such as legal ontologies in Web Ontology Language (OWL)². This is often done by transforming the source legal documents into XML standards and tagging the relevant information to allow queries and information retrieval from the XML files.

However, the overall usefulness of these systems are limited due to their focus on terminological issues and information retrieval while disregarding the specific semantic aspects, which allow for legal reasoning. Just as standard deontic logic mostly focused on the notion of obligation, subsequent developments in deontic logic also adopted an abstract view of law, with a very loose connection with the texts of regulations. For lawyers, the meaning of laws can be fully understood only within the rich expressiveness of natural language since “like language generally, legal discourse can never escape its own textuality” [47].

There is thus a gap between a powerful reasoning mechanism on the formalization of law and the textuality of law, which can be addressed with solutions coming from the literature on Natural Language Semantics.

A new standardization initiative called LegalRuleML³ [13; 14] tries to address

²See <https://www.w3.org/OWL>

³See <https://www.oasis-open.org/committees/legalruleml>

these issues. LegalRuleML is an XML format that extends the RuleML standard⁴ to define a rule interchange language for the legal domain. While legal XML standards are used to tag the original textual content of the legal documents, LegalRuleML separately represents and stores the logical content of the provisions. Specifically, LegalRuleML allows to specify semantic/logical representations and associate them with both the structural elements of the documents or with tasks in a business process. LegalRuleML allows to encode RuleML representations of formulas⁵ in Defeasible Deontic Logic (DDL) [30]. This is an extension of standard Defeasible Logic with deontic operators, and the operators for compensatory obligation [34]. Defeasible Logic is an efficient and simple rule based computationally oriented non-monotonic formalism, designed for handling exceptions in a natural way. According to the formalization proposed in [12], Defeasible Logic is a constructive logic with its proof theory and inference condition as its core. The logic exploits both positive proofs, where a conclusion has been constructively proved using the given rules and inference conditions (also called proof conditions), and negative proofs: showing a constructive and systematic failure of reaching particular conclusions, or in other terms, constructive refutations. The logic uses a simple language, that proved to be successful in many application area, due to its scalability and constructiveness. These elements are extremely important for normative reasoning, where an answer to a verdict is often not enough, and full traceability is needed.

2.3 Legal reasoning and Defeasible Deontic Logic

Norms describe general cases and what behavior should be taken, or the consequences, if the real facts are similar to the general case described in the norm. Therefore, norms describe the conditions under which they are applicable and the normative effects they produce when applied. Simply put, the scope of norms is to regulate the behavior of their subjects and to define what is legal and what is illegal.

In a compliance perspective, the normative effects of importance are the deontic effects (also called normative positions). The basic and more important deontic effects are: obligation, prohibition and permission.

- **Obligation:** when there is a situation, an act, or a course of action to which a bearer is legally bound, and if it is not achieved or performed results in a

⁴See <http://wiki.ruleml.org>

⁵However, LegalRuleML is actually logic-neutral, i.e., it permits to encode formulae in other logics, even radically different from Defeasible Deontic Logic. For instance, [46] and [49] presents an ontology and a knowledge base of formulae that formalizes the norms in the GDPR. This knowledge base will be possibly considered in future works, because at present there is not a reasoner such as Regorous that works with reified I/O logic formulae.

violation.

- **Prohibition:** when there is a situation, an act, or a course of action which a bearer should avoid, and if it is achieved results in a violation.
- **Permission:** when something is permitted if the prohibition of it or the obligation to the contrary do not hold.

This gives rise to some considerations:

- Obligations and prohibitions are constraints that limit the space of action of processes.
- They can be violated, and a violation does not imply an inconsistency within a process with the consequent termination of or impossibility to continue the business process.
- Violations can be generally compensated for, and processes with compensated violations are still compliant [35; 32] (e.g. contracts typically contain compensatory clauses specifying penalties and other sanctions triggered by breaches of other contracts' clauses [27]).
- Not all violations are compensable, and uncompensated violations means that a process is not compliant.
- Permissions cannot be violated. They can be used (indirectly) to determine that there are no obligations or prohibitions to the contrary, or to derive other deontic effects.
- Legal reasoning and legal theory typically assume a strong relationship between obligations and prohibitions: the prohibition of A is the obligation of $\neg A$ (the opposite of A), and then if A is obligatory, then $\neg A$ is forbidden [53].

Taking in consideration the notion of obligation, compliance means to identify whether a process violates or not a set of obligations. Thus, the first step is to determine whether and when an obligation is in force. Hence, an important aspect of the study of obligations is to understand the lifespan of an obligation and its implications on the activities carried out in a process. A norm can specify if there is:

- **Punctual obligations:** an obligation is in force for a particular time point.

- **Persistent obligations:** a norm indicates when an obligation enters in force. An obligation remains in force until terminated or removed.
 - For persistent obligations we can ask if to fulfil an obligation we have to obey to it for all instants in the interval in which it is in force, **maintenance obligations**, or
 - Whether doing or achieving the content of the obligation at least once is enough to fulfil it, **achievement obligations**.
 - For achievement obligations another aspect to consider is whether the obligation could be fulfilled even before the obligation is actually in force. If this is admitted, then there is a **preemptive obligation**, otherwise the obligation is **non-preemptive**.
- **Termination of obligations:** norms can specify the interval in which an obligation is in force.

As said, what differentiates obligations from other constraints is that obligations can be violated.

- If we still have to comply with a violated obligation (the obligation persists after being violated) we speak of a **perdurant obligation**.
- Otherwise, we speak of a **non-perdurant obligation** [28].

3 The project CANP

Our work is collocated within CANP project⁶, which aims at using Artificial Intelligence to enhance e-Health procedures within the Città della Salute e della Scienza di Torino⁷, the biggest hospital complex in Europe [55]. A case study of the project is concerned with the application of innovative telemedicine technologies supporting the care of elderly patients in the context of a Hospital at Home (HaH). The use of communication systems in the remote management of the patient could improve treatment outcomes, increase access to care, and reduce health costs [24].

We show below how it is possible to model and integrate, within the HaH process, compliance checking via DDL and the Regorous reasoner. As mentioned above, we consider GDPR provisions to safeguard the personal data of the patients, but the approach is general enough to handle any kind of legal constraint involved in e-Health procedures.

⁶<http://casanelparco-project.it>

⁷<https://www.cittadellasalute.to.it>

3.1 Hospital at home (HaH)

For more than 30 years, the "Città della Salute e della Scienza of Turin has operated the Hospital at Home (HaH). This is a home care service defined by Resolution DGR n. 85-13580 of 16 March 2010, as a form of health care hospital character, which provides for the organization of care in the home of patients suffering from acute diseases, but who do not require equipments with high technological complexity and intensive or invasive monitoring [54].

The service is composed by two main processes: the acceptance (in Fig. 1) and the tour visits in the patients' houses (in Fig. 2)⁸.

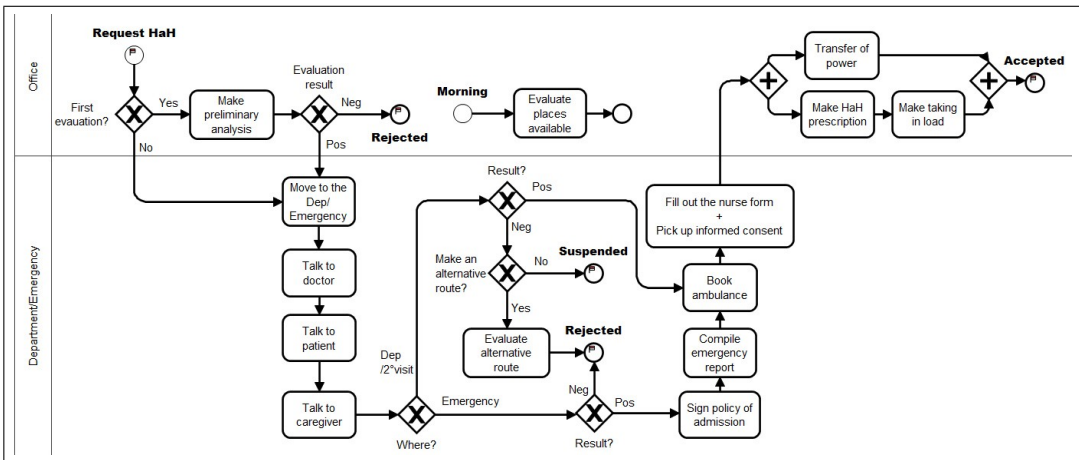


Figure 1: The patient registration process model of HaH service in standard language BPMN.

Requests for the activation of the HaH service are made by the emergency or regular departments and by general medical doctors. After that, each patient is evaluated by the team to establish the feasibility of hospitalization under HaH.

The service begins with the admission process shown in Fig. 1. It involves the Case Manager (CM), who has to evaluate all the requests. Each case refers to some guidelines to understand if the patient has some characteristics to take in charge to this type of hospitalization. At the end of this evaluation process, for the taking in charge of a patient, a real contract of collaboration is created.

The contract involves on one side the hospital, and in particular the staff of the department of HaH, and on the other side the patient with the caregiver and possibly

⁸For reasons of space in this article are illustrated only the salient features of the processes, for a more accurate description see [3; 10; 11]

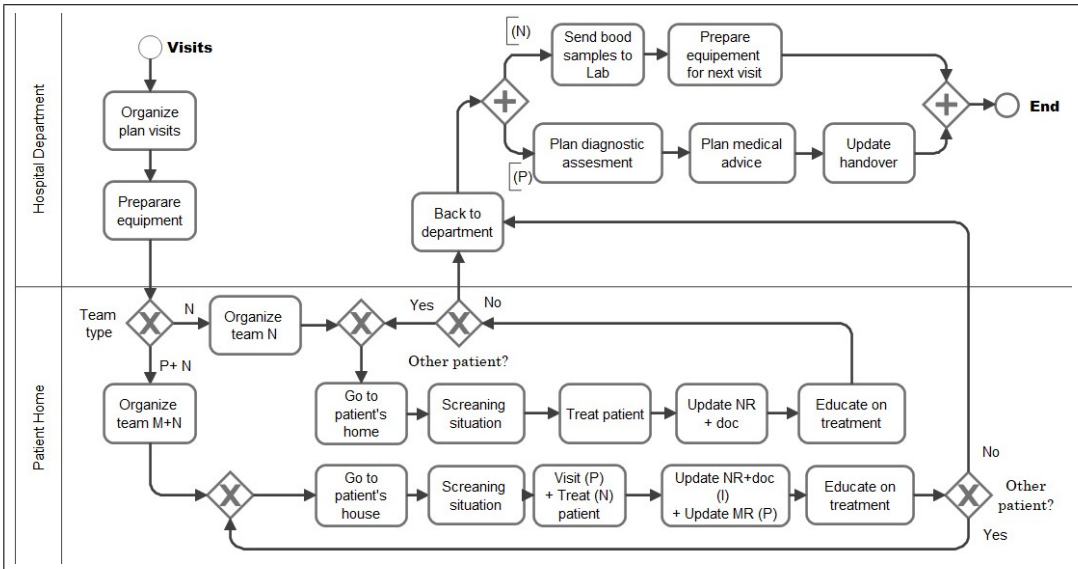


Figure 2: Hospital at home process model including Hospital Department and Patient Home lanes in BPMN

with his/her family, which can coincide with the figure of the caregiver. Besides, it is important that this type of collaboration remains as initially established for the whole duration of the service. Otherwise, for instance, in case of missing caregiver or family exhaustion, the patient is immediately move to the hospital and hospitalized classically inside the hospital wall.

Firstly, the CM has to evaluate every morning the available number of possible posts (**Evaluation n° places available**), that correspond to the maximum number of patients that she could accept in this day. She has to evaluate the probable number of discharged patient, the available staff, how long each patients, they already have in charge, been (some patients have some pathology that requires more time than others, for example blood transfusions are longer than bandages that are longer than giving a medicine. The first type of patients occupies two slots, the second type of patients occupies one slot and an half and the third type occupies only one place).

This first evaluation determines the future workload of all the staff involved in the service. At the same time, input requests can arrive by telephone from the emergency department as well as from any other hospital department. The requests are made by the responsible doctors of the departments that made a first quick evaluation.

The arrival of a request by phone at the Hospital at Home (HaH) (generator **Re-**

quest HaH) implies an initial evaluation (gateway *First evaluation?*) by the doctor and the CM or the chief nursing (**Make preliminary analysis**). If there are features not complying with this type of hospitalization (gateway *Evaluation's result?*) the request is immediately rejected (end of the process **Rejected**). Otherwise, CM moves to the department to evaluate the patient (**Move to the Dep/Emergency**). At first, the CM talks to the requested doctor to evaluate clinical conditions (**Talk to doctor**). All patients are in acute disease but they must not be in state of bleeding or risk of reanimation. Then the CM talks to the patient to check if he/she is conscious and capable of understanding and willing (**Talk to patient**), as well as to the family and the caregiver (**Talk to caregiver**).

During this meeting the CM explains to the patient, if possible, and to the family the characteristics, organization and requirements of the service. On the other hand, she evaluates clinical, functional and cognitive aspects.

Through this structured interview of mutual knowledge, the CM attentively appraises the real availability to accept the cares in house, if it is possible to identify a caregiver, so the availability of taking in charge the patient in this type of hospitalization.

The requests could be forwarded both from each department of the hospital and from the emergency department. For both of them the activities already shown are always the same, but after having talked to all the interested parts, the decisional trial is different according to where they are (gateway *Where?*).

If they are in the emergency department there is an urgent need to free up beds. Any bed of the emergency department can be busy for more than 24 hours. Therefore, the evaluation result must be immediately positive or negative (gateway *Result?*). If it is negative the request is definitively rejected (**Rejected**). Probably the patient has not the requirement and he is transferred in a standard department. If the parts (CM-patient-caregiver-patient's family) reach the accord to hospitalize at home the CM signs the policy of admission (**Sign policy of admission**), the emergency department's doctor compiles the emergency report (**Compile emergency report**) and then the CM books the ambulance for the transport to the patient's domicile with the transport settled with the hospital (**Book ambulance**) and finally the CM fills out the nurse form asking dates to the patient/caregivers, collects some patient's information, gives to the patient and his/her family some information about the service including an "Informative Card" with information on the service and about organization of the next tasks, and at the end makes to sign and pick up informed consent to the patient, or to the caregiver if the patient is unable (**Fill out the nurse form + Pick up informed consent**).

If the request came from a standard department of the hospital the result of the evaluation (gateway *Result?*) could be:

- Positive: the patient is taken in charge, so the CM books the ambulance, gives and takes different information, fill out the nurse form and make sign the informed consent to the patient, like the previous process (**Book ambulance** and **Fill out the nurse form + Pick up informed consent**).
- Really negative: the CM suggests an alternative route to the patient (gateway *Make an alternative route?*) and the request for this type of hospitalization is definitively rejected (**Rejected**).
- Negative but actually Suspended: often the family needs time to organize themselves or to require medical products or it is necessary to talk also to the “real” caregiver that will actually stay with the patient or to other family members, so it is a temporary rejection (**Suspended**), but the CM takes another appointment.

To establish this contract of trust and collaboration among patient and hospital, it is essential that the CM talks to the whole family nucleus to establish a closer contact with the patient, that must take care and divide assignments and responsibility and finally with the caregiver, who might also be a relative or not. It is necessary that all these people are informed, aware and give the consent to the service, otherwise there could be severe consequences in terms of collaboration that could affect the patient’s care.

In this case, the CM will have other tours (gateway *First evaluation* arrow 2° visit). These others visit are in average 1, 2, 3 or at most 4 in particular cases (e.g., if there is the need to wait some medical products that have to be ordered). These other visits are not made by a different doctor with other requests, but the CM takes the appointment on a case-by-case bases directly with the patients. The activities remain the same but need less time than the firsts. This second evaluation could exist only in the department (gateway *Where*, 2° visit), as has been already explained. In all these visits, it is possible both a taking in charge of the patient, or a rejection of the request, or a suspension of the request which will generate another visit, and the trial can be repeated until the patient will be taken in charge, or the service will be refused, or the patient will die or will be discharged.

In all cases in which the patient go at home in a different day from that of the request of the HaH, the CM autonomously goes to the patient before he goes away, with the purpose to make sure that all the information are clear. It imply the remake of the three activities already explain but in less time.

At the end of this trial with the patient the CM comes back to her department’s office and makes the administrative tasks for the patients just taken in charge. On the hospital’s computer system the CM has to make the prescription of the

Hospitalization at Home (**Make HaH prescription**) and the formal taking in charge in the department of the Hospitalization at Home (**Make taking in charge**). In the meantime, as soon as the doctors and the nurses arrive the CM informs them about the new patients (**Transfer of power**). At this time the request is also formally accepted and the patient is definitively in the workload of the department of the Hospital at Home (**Accepted**).

Fig. 2 shows the business process of the Hospital at Home service, in detail the organization of the tour visits of the staff (medical doctors and nurses) going to patient's home.

All patients receive home visits every morning; some patients with special conditions (politransfused or antibiotic therapy) may also receive an afternoon visit.

At full workload, there are 7 nurses and 4 physicians in the morning, and there are 2 nurses and 1 physician in the afternoon involved in the shown process. This staff is then divided into teams to carry out tours. In the morning there are 6 teams: 4 teams composed of 1 physician (or 1 grad student) + 1 nurse and 2 teams made by 1 nurse. In the afternoon there are 2 teams: 1 team made by 1 physician + 1 nurse and 1 team of 1 nurse. Each team visits on average 4 patients.

In the morning, all the staff together analyze all the patient's situations according to four impact factors: medical and nursing complexity care, condition of the caregiver and geographical location of the house's patient (**Organize tour visits**). This allows to divide the whole amount of patients in balanced groups in terms of time to spend in visits and time to go from patient to patient; and assign to each group of patients an hospital team (gateway *Team type?* composed by one physician + one nurse, **Organize team PN**, or made by only one nurse, **Organize team N**). After that, each nurse prepares the medical equipment for each of his patients (**Prepare equipment**).

Once arrived at the patient's home (**Move to home patient**), they analyze the current situation (**Screening situation**) and carry out the visit. If there is only the nurse, he treats the patient (**Treat patient**), updates both the nurse record and the other clinical and organizational documents (**Update NR + doc**) and educates the caregiver on treatment (**Educate on treatment**). If there are both physician and nurse, the physician visits the patient while the nurse treats him (**Visit (P) + Treat (N) patient**); after, the physician updates the medical record and the nurse compiles the nurse record and the other organizational documents (**Update NR+doc (N) + Update MR (P)**) and, at the end, they educate the caregiver on treatment (**Educate on treatment**).

Once the visit is finished, if there is another patient to visit (gateway *Another patient?*), the team heads to the second patient's house. The cycle resumes until the assigned patients are not finished; only then the team will be back to the hospital

(Back to hospital).

The morning shift staff completes some “administrative” tasks (**Plan diagnostic assessment, Plan medical advice, Update handover** for physicians and **Send blood samples to Laboratory, Prepare equipment for next visits** for nurses)

In the meanwhile, all the staff make the handover: the morning staff communicates the different patients’ situation, one by one; and the afternoon staff receives any useful information to organize the future work. Subsequently, they **organize tour visits**, the nurses **Prepare equipment**, they decide the team composition and start the visit tour. All the activities are the same already explained for the morning.

3.2 Compliance check in HaH

Compliance checking not only refers to the tasks that an organization must perform to achieve its business goals, but also to their effects, i.e., how the activities in the tasks change the environment in which they operate, and the artefacts produced by the tasks (see discussion in [37]). To capture these aspects, process models are usually enriched with semantics annotations [52]. Each task in a process model can have attached to it a set of semantic annotations. Annotations are formal representations, e.g., formulae, giving a description of the environment in which a process operates. Then, it is possible to associate with each task in a trace a set of formulas corresponding to the state of the environment after the task has been executed in that particular trace. It is important to underline that different traces can result in different states, even if the tasks in the traces are the same. Moreover, even if the end states are the same, the intermediate states can be different. Finally, a trace uniquely determines the sequence of states obtained by executing the trace.

The business compliance checking tool Regorous [29] allows to enrich BPMN graphs with semantics annotations corresponding to DDL formulas.

As part of our research activity in the “CANP” project, we have enriched the BPMN graphs representing e-Health processes within the “Città della Salute e della Scienza” of Turin, such as the one shown above in Fig. 1, with selected GDPR legal constraints modeled in DDL and LegalRuleML. These constraints have been then implemented in Regorous in order to test and evaluate compliance checking with respect to different input configurations and scenarios. The LegalRuleML and Regorous formalizations of the GDPR norms that we considered are available online⁹.

Regorous implements the sub-classes of obligations and permission seen above in the section “Legal reasoning and Defeasible Deontic Logic” via the following notations (atomic DDL formulas):

⁹See <https://github.com/liviorobaldo/BPMinHealthcare>

- **[P]p**: p is permitted.
- **[OM]p**: there is a maintenance obligation for p.
- **[OAPP]p**: there is an achievement preemptive and perdurant obligation for p.
- **[OAPNP]p**: there is an achievement preemptive and non-perdurant obligation for p.
- **[OANPP]p**: there is an achievement non preemptive and perdurant obligation for p.
- **[OANPNP]p**: there is an achievement non preemptive and non-perdurant obligation for p.

In the above notations, “p” is a predicate, called a “term” in Regorous terminology. Regorous lists all terms used in a set of formalizations, together with their description, in a special XML tag `<vocabulary>`. Two terms used in the formalization of Art. 6 of GDPR are the following:

```
<vocabulary>
  <Term atom="Proc" description="Processing: means any
    operation or set of operations which is performed on
    personal data ..."/>
  <Term atom="GiveConsent" description="Consent given
    by the data subject means any freely given, specific,
    informed and unambiguous indication ..."/>
</vocabulary>
```

On the other hand, (part of) the formalization of Art. 6 is the following; note that we chose to formalize the processing of personal data as prohibited unless one of the legal basis is in place, e.g., unless the patient has given consent to the processing of personal data (see GDPR, Art.6.1(a)):

```
<Rule xmlns:xsi="..." xsi:type="DflRuleType" ruleLabel="Art.6.0">
  <ControlObjective>Personal data processing
    is prohibited.</ControlObjective>
  <FormalRepresentation>=>[OM]-Proc</FormalRepresentation>
</Rule>
```

```
<Rule xmlns:xsi="..." xsi:type="DflRuleType" ruleLabel="Art.6.1a">
  <ControlObjective>Processing shall be lawful if the data
    subject has given consent to the processing of his or
    her personal data for one or more specific
    purposes.</ControlObjective>
  <FormalRepresentation>GiveConsent=>[P]Proc</FormalRepresentation>
</Rule>
```

“-” and “=>” are the standard propositional logic operators for negation and implication. Thus, the two formulas above can be rewritten in a more classical notation as “=>[OM]-Proc” and “GiveConsent=>[P]Proc”.

Of course, the two formulas cannot hold together as the first entails that the processing is prohibited while the latter entails that it is permitted. In order to solve these conflicts, both LegalRuleML and DDL implement overriding relations between norms. In our example, the second formula will have to override the first one, in order to permit processing of personal data when consent is given.

In Regorous, overriding is implemented as “superiority relations”, encoded via the homonym tag, in which the “superiorRuleLabel” overrides the “inferiorRuleLabel”. In the example under consideration we have:

```
<SuperiorityRelation superiorRuleLabel="Art.6.1a"
  inferiorRuleLabel="Art.6.0"/>
```

Given a set of well-formed rules and superiority relations encoded in the XML format briefly seen above, Regorous allows to check whether a Business Process in the BPMN standard is compliant with them.

Regorous is implemented as a plug-in of Eclipse¹⁰. The BPMN is uploaded in the platform together with a set of rules in Regorous XML format. Subsequently, in each task of the process it is possible to specify which terms of the vocabulary are true or false via special Eclipse windows provided by the plug-in. Of course, the truth value of these terms might be also asserted programmatically during the real-time execution of the Business Process; this is indeed how we plan to use Regorous in the future, when the service will be up and running. However, since at present we are still in the research and development phase, in our current activity we always executed Regorous by manually identifying, setting, and testing different input configurations and scenarios.

Fig.3 shows a simple example of how Regorous performs compliance checking on the BPMN representation from Fig.1. The BPMN file is uploaded in Eclipse together with the ruleset formalizing the GDPR norms in Regorous XML format.

¹⁰<https://www.eclipse.org>

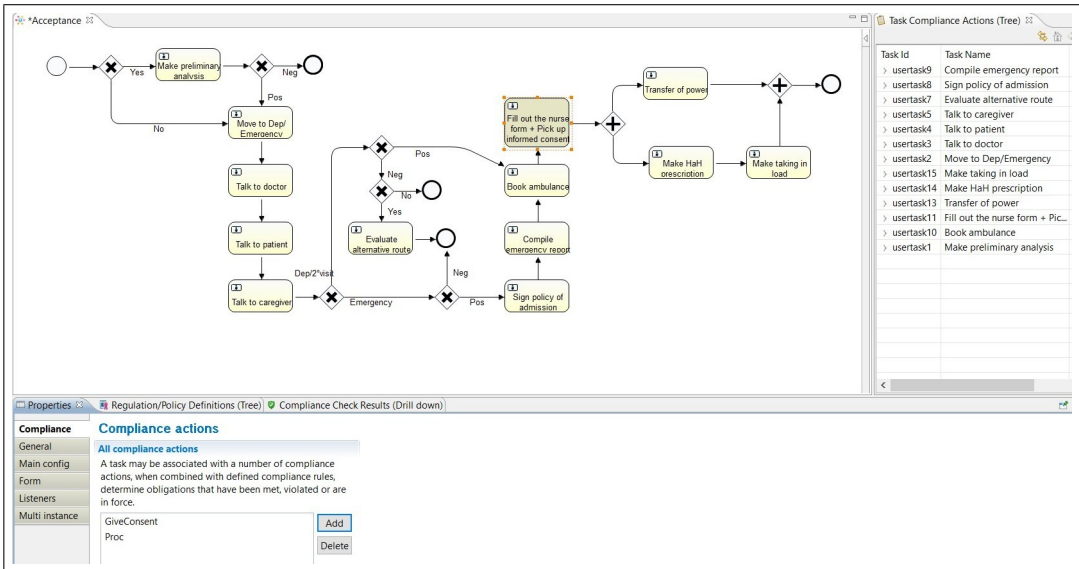


Figure 3: Regorous screenshot example for compliance checking.

The plug-in includes special tabs (shown on the bottom of Fig. 3) that allow to specify, for each task, the values of the terms. For instance, by specifying “GiveConsent” and “Proc” in the task “Fill out the nurse form + Pick up informed consent”, Regorous infers that the process is compliant with the ruleset, as the superiority relation seen above will make the processing of personal data permitted. Conversely, by specifying the single action “Proc”, Regorous infers that the process is not compliant with the ruleset because the rule with ruleLabel="Art.6.0" asserts the processing of personal data as prohibited and, contrary to the previous case, that prohibition is not overridden by a stronger permission.

After specifying the rules or checks performed in that task in the various activities, it is possible to run the Regorous check. Thank to the superiority rules and the BPMN, for the check Regorous will follow the flow of the process, in this way it is able not only to check if every rule is respected, but also if the sequence of them is compliant to the sequence imposed by law.

If the result of the compliance checking is positive it will appear a green screen as in Fig. 4.

if the control detects non-conformities or anomalies, the same screen will appear but red or orange respectively, which will highlight in which areas the non-conformities were detected and with respect to which controls.

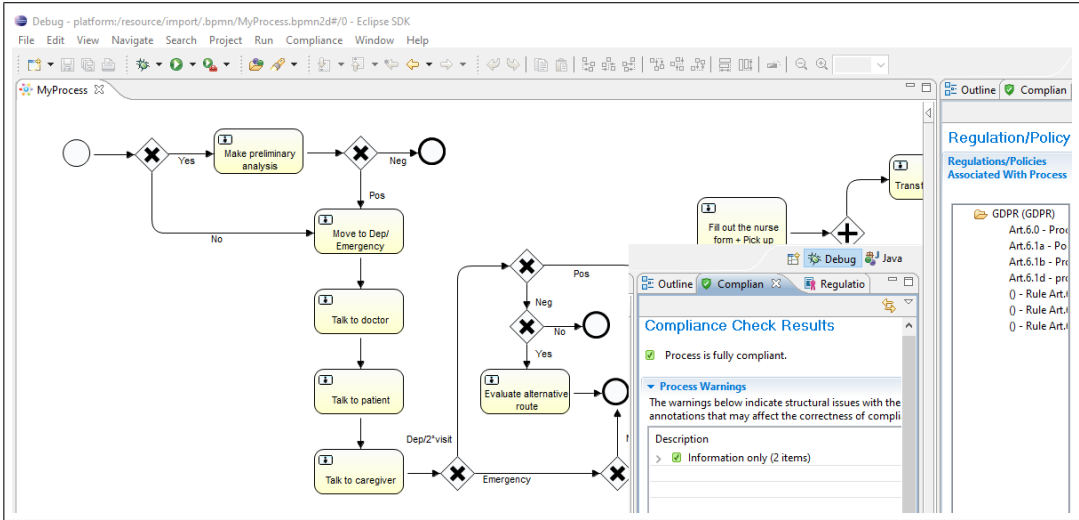


Figure 4: Regorous screen of compliance check results.

4 Conclusion and Future Works

The proposed pipeline addresses a specific risk management application in a selected healthcare process. However, of course further work is needed to formalize all the laws, regulations, and guidelines involved in all healthcare processes, in order to have a full exhaustive analysis on how legal compliance is handled in medical procedures. Although, as explained in Section 2, the DDL is currently one of the best logics for formalising legal rules, this formalisation still needs to be done by a legal expert, who has experience with the principles governing legal interpretation, both for formalize the rules and for establish the relations of superiority between them.

On the other hand, using this methodology, if changes are made over time, you can only change the impacted area, leaving the rest of the work intact:

- If the business process is changed, just modify the modified activities in the real process to check if the new process is still compliant (without changing the rest of the activities).
- Since the law is formalized in an XML file and then uploaded to Regorous:
 - If the legislation is changed, the XML file can be modified only in the parts modified by the legislator (without changing the rest of the corpus).
 - If we have a second process that must comply with a regulation already formalized, just add the XML file to the second BPMN (without having

to remake the formalization already made for the first process).

- If a new law is added in the field of our business, it will be enough to add a second XML file containing the new legislation. In this way, the compliance check will be carried out for both regulations (without changing the BPMN or the previous XML files).

In conclusion, the aim is to combine this compliance checking methodology in a context of re-organization and optimization of processes. Maintaining the already formalized norms as a background, the purpose is to obtain a methodology able to balance the managerial aspect with that of regulatory compliance.

Finally, the authors are also currently working on two branches of research. On one hand, in the context of the project “CANP”, on the development of a methodology to automatize or semi-automatize formalization of laws, that combines Defeasible Deontic Logic with NLP technologies, in order to make the whole process faster, simpler, and accessible to users who have little or no competence in law or in logical formalizations. On the other hand, the authors are working on the automation of the compliance checking process starting from a legal point of view, i.e., by seeking methodologies capable of reproducing the principles governing legal interpretation [4; 57; 58].

Acknowledgments

The research presented in this paper received funding from the project “CANP” (<http://casanelparco-project.it>), supported by European Regional Development Fund (ERDF) 2014/2020; European Social Fund (ESF) 2014/2020. Livio Robaldo has been supported by the Legal Innovation Lab Wales operation within Swansea University’s HRC School of Law; the operation has been part-funded by the European Regional Development Fund through the Welsh Government.

References

- [1] Waleed Abo-Hamad and Amr Arisha. Simulation-based framework to improve patient experience in an emergency department. *European Journal of Operational Research*, 224(1):154–166, 2013.
- [2] Thomas Allweyer. *BPMN 2.0: introduction to the standard for business process modeling*. Books on Demand, 2016.
- [3] Ilaria Angela Amantea, Marzia Arnone, Antonio Di Leva, Emilio Sulis, Dario Bianca, Enrico Brunetti, and Renata Marinello. Modeling and simulation of the hospital-at-home service admission process. In *SIMULTECH*, pages 293–300, 2019.

- [4] Iliaria Angela Amantea, Luigi Di Caro, Llio Humphreys, Rohan Nanda, and Emilio Sulis. Modelling norm types and their inter-relationships in eu directives. In *ASAIL@ ICAIL*, 2019.
- [5] Iliaria Angela Amantea, Antonio Di Leva, and Emilio Sulis. A simulation-driven approach in risk-aware business process management: A case study in healthcare. In *SIMULTECH*, pages 98–105, 2018.
- [6] Iliaria Angela Amantea, Antonio Di Leva, and Emilio Sulis. A simulation-driven approach to decision support in process reorganization: a case study in healthcare. In *Proceedings of the 15th Conference of the Italian Chapter of AIS*, volume 1, pages 98–105. ITAIS, 2018.
- [7] Iliaria Angela Amantea, Antonio Di Leva, and Emilio Sulis. Risk-aware business process management: a case study in healthcare. In *The Future of Risk Management, Volume I*, pages 157–174. Springer, 2019.
- [8] Iliaria Angela Amantea, Antonio Di Leva, and Emilio Sulis. A simulation-driven approach to decision support in process reorganization: A case study in healthcare. In *Exploring Digital Ecosystems*, pages 223–235. Springer, 2020.
- [9] Iliaria Angela Amantea, Antonio Di Leva, and Emilio Sulis. A simulation-driven approach in risk-aware business process management: A case study in healthcare. In *Proceedings of 8th International Conference on Simulation and Modeling Methodologies, Technologies and Applications - Volume 1: SIMULTECH*,, pages 98–105. INSTICC, SciTePress, 2018.
- [10] Iliaria Angela Amantea, Emilio Sulis, Guido Boella, Andrea Crespo, Dario Bianca, Enrico Brunetti, Renata Marinello, Marco Grosso, Jan-Christoph Zoels, Michele Visciola, et al. Adopting technological devices in hospital at home: A modelling and simulation perspective. In *SIMULTECH*, pages 110–119, 2020.
- [11] Iliaria Angela Amantea, Emilio Sulis, Guido Boella, Renata Marinello, Marco Grosso, and Andrea Crespo. A modeling framework for an innovative e-health service: The hospital at home. In *International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, pages 111–132. Springer, 2020.
- [12] Grigoris Antoniou, David Billington, Guido Governatori, and Michael J Maher. Representation results for defeasible logic. *ACM Transactions on Computational Logic (TOCL)*, 2(2):255–287, 2001.
- [13] Tara Athan, Harold Boley, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. Legalruleml: From metamodel to use cases. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 13–18. Springer, 2013.
- [14] Tara Athan, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. Legalruleml: Design principles and foundations. In *Reasoning Web International Summer School*, pages 151–188. Springer, 2015.
- [15] Cesare Bartolini, Andra Giurgiu, Gabriele Lenzi, and Livio Robaldo. Towards legal compliance by correlating standards and laws with a semi-automated methodology. In *BNCAI*, volume 765 of *Communications in Computer and Information Science*, pages

- 47–62. Springer, 2016.
- [16] Jörg Becker, Patrick Delfmann, Mathias Eggert, and Sebastian Schwittay. Generalizability and applicability of model-based business process compliance-checking approaches—a state-of-the-art analysis and research roadmap. *Business Research*, 5(2):221–247, 2012.
 - [17] Stefanie Betz, Susan Hickl, and Andreas Oberweis. Risk-aware business process modeling and simulation using xml nets. In *Commerce and enterprise computing (cec), 2011 IEEE 13th conference on*, pages 349–356. IEEE, 2011.
 - [18] Guido Boella, Luigi Di Caro, Llio Humphreys, Livio Robaldo, Piercarlo Rossi, and Leendert van der Torre. Eunomos, a legal document and knowledge management system for the web to provide relevant, reliable and up-to-date information on the law. *Artificial Intelligence and Law*, 24(3):245–283, 2016.
 - [19] Guido Boella, Luigi Di Caro, Llio Humphreys, Livio Robaldo, and Leon van der Torre. Nlp challenges for eunomos, a tool to build and manage legal knowledge. *Language resources and evaluation (LREC)*, pages 3672–3678, 2012.
 - [20] Guido Boella, Luigi Di Caro, Daniele Rispoli, and Livio Robaldo. A system for classifying multi-label text into eurovoc. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, ICAIL '13*, pages 239–240, New York, NY, USA, 2013. ACM.
 - [21] Guido Boella, Luigi Di Caro, Daniele Rispoli, and Livio Robaldo. A system for classifying multi-label text into eurovoc. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, pages 239–240, 2013.
 - [22] Guido Boella, Luigi Di Caro, and Livio Robaldo. Semantic relation extraction from legislative text using generalized syntactic dependencies and support vector machines. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 218–225. Springer, 2013.
 - [23] James Joseph Buddle, Brenda Sue Burke, Rodney Albert Perkins, Leon Ellis Roday, Renee Tartaglia, and Ivan Antonio Vermiglio. System and method for compliance management, 2005. US Patent 6,912,502.
 - [24] Gideon A Caplan, Nur S Sulaiman, Dee A Mangin, Nicoletta Aimonino Ricauda, Andrew D Wilson, and Louise Barclay. A meta-analysis of “hospital in the home”. *Medical Journal of Australia*, 197(9):512–519, 2012.
 - [25] Antonio Di Leva, Emilio Sulis, Angela De Lellis, and Ilaria Angela Amantea. Business process analysis and change management: The role of material resource planning and discrete-event simulation. In *Exploring Digital Ecosystems*, pages 211–221. Springer, 2020.
 - [26] M. Dumas, M. La Rosa, J. Mendling, and H. Reijers. *Fundamentals of business process management*, volume 1. Springer, 2nd edition, 2018.
 - [27] Guido Governatori. Representing business contracts in ruleml. *International Journal of Cooperative Information Systems*, 14(02n03):181–216, 2005.
 - [28] Guido Governatori. The regorous approach to process compliance. In *2015 IEEE 19th*

- International Enterprise Distributed Object Computing Workshop*, pages 33–40. IEEE, 2015.
- [29] Guido Governatori. Thou shalt is not you will. In *Proceedings of the 15th international conference on artificial intelligence and law*, pages 63–68, 2015.
- [30] Guido Governatori. Practical normative reasoning with defeasible deontic logic. In *Reasoning Web International Summer School*, pages 1–25. Springer, 2018.
- [31] Guido Governatori and Mustafa Hashmi. No time for compliance. In *2015 IEEE 19th International Enterprise Distributed Object Computing Conference*, pages 9–18. IEEE, 2015.
- [32] Guido Governatori and Zoran Milosevic. Dealing with contract violations: formalism and domain specific language. In *Ninth IEEE International EDOC Enterprise Computing Conference (EDOC'05)*, pages 46–57. IEEE, 2005.
- [33] Guido Governatori, Zoran Milosevic, and Shazia Sadiq. Compliance checking between business processes and business contracts. In *2006 10th IEEE International Enterprise Distributed Object Computing Conference (EDOC'06)*, pages 221–232. IEEE, 2006.
- [34] Guido Governatori and Antonino Rotolo. Logic of violations: A gentzen system for reasoning with contrary-to-duty obligations. *The Australasian Journal of Logic*, 4, 2006.
- [35] Guido Governatori and Shazia Sadiq. The journey to business process compliance. In *Handbook of research on business process modeling*, pages 426–454. IGI Global, 2009.
- [36] Yacov Y Haimen. *Risk modeling, assessment, and management*. John Wiley & Sons, 2015.
- [37] Mustafa Hashmi, Guido Governatori, and Moe Thandar Wynn. Business process data compliance. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 32–46. Springer, 2012.
- [38] Mustafa Hashmi, Guido Governatori, and Moe Thandar Wynn. Modeling obligations with event-calculus. In *International Symposium on Rules and Rule Markup Languages for the Semantic Web*, pages 296–310. Springer, 2014.
- [39] Mustafa Hashmi, Guido Governatori, and Moe Thandar Wynn. Normative requirements for regulatory compliance: An abstract formal framework. *Information Systems Frontiers*, 18(3):429–455, 2016.
- [40] John Hayes. *The theory and practice of change management*. Palgrave Macmillan, 2014.
- [41] Linh Thao Ly, Fabrizio Maria Maggi, Marco Montali, Stefanie Rinderle-Ma, and Wil MP van der Aalst. Compliance monitoring in business processes: Functionalities, application, and tool-support. *Information systems*, 54:209–234, 2015.
- [42] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: Concepts, techniques and tools*. Princeton university press, 2015.
- [43] Richard Müller and Andreas Rogge-Solti. Bpmn for healthcare processes. In *Proceedings of the 3rd Central-European Workshop on Services and their Composition (ZEUS 2011), Karlsruhe, Germany*, volume 1, 2011.
- [44] Jorge Munoz-Gama, Niels Martin, Carlos Fernandez-Llatas, Owen A Johnson, Marcos

- Sepúlveda, Emmanuel Helm, Victor Galvez-Yanjari, Eric Rojas, Antonio Martinez-Millana, Davide Aloini, et al. Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, 127:103994, 2022.
- [45] Rohan Nanda, Luigi Di Caro, Guido Boella, Hristo Konstantinov, Tenyo Tyankov, Daniel Traykov, Hristo Hristov, Francesco Costamagna, Llio Humphreys, Livio Robaldo, and Michele Romano. A unifying similarity measure for automated identification of national implementations of european union directives. In Jeroen Keppens and Guido Governatori, editors, *Proc. of the 16th edition of the International Conference on Artificial Intelligence and Law, ICAIL 2017*. ACM, 2017.
- [46] Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. Pronto: Privacy ontology for legal compliance. In *Proceedings of the 18th European Conference on Digital Government (ECDG)*, October 2018.
- [47] Gary Peller. The metaphysics of american law. *Calif. L. Rev.*, 73:1151, 1985.
- [48] Nicolas Racz, Edgar Weippl, and Andreas Seufert. A process model for integrated it governance, risk, and compliance management. In *Proceedings of the Ninth Baltic Conference on Databases and Information Systems (DB&IS 2010)*, pages 155–170. Citeseer, 2010.
- [49] L. Robaldo, C. Bartolini, M. Palmirani, A. Rossi, M. Martoni, and G. Lenzini. Formalizing gdpr provisions in reified i/o logic: the dapreco knowledge base. *The Journal of Logic, Language, and Information*, 29, 2020.
- [50] Stefan Sackmann, Stephan Kühnel, and Tobias Seyffarth. Using business process compliance approaches for compliance management with regard to digitization: Evidence from a systematic literature review. 09 2018.
- [51] Kit Sadgrove. *The complete guide to business risk management*. Routledge, 2016.
- [52] Shazia Sadiq, Guido Governatori, and Kioumars Namiri. Modeling control objectives for business process compliance. In *International conference on business process management*, pages 149–164. Springer, 2007.
- [53] Giovanni Sartor. Legal reasoning. *A Treatise of Legal Philosophy and General Jurisprudence*, 5, 2005.
- [54] Emilio Sulis, Iliaria Angela Amantea, Guido Boella, Renata Marinello, Dario Bianca, Enrico Brunetti, Mario Bo, Alessandra Bianco, Francesco Cattel, Clara Cena, et al. Monitoring patients with fragilities in the context of de-hospitalization services: an ambient assisted living healthcare framework for e-health applications. In *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*, pages 216–219. IEEE, 2019.
- [55] Emilio Sulis, Clara Cena, Roberta Fruttero, Sara Traina, Luca Carlo Feletti, Pierluigi de Cosmo, Lucrezia Armando, Serena Ambrosini, Iliaria Angela Amantea, Guido Boella, Renata Marinello, Dario Bianca, Enrico Brunetti, Mario Bo, Alessandra Bianco, and Francesco Cattel. Monitoring patients with fragilities in the context of de-hospitalization services: An ambient assisted living healthcare framework for e-health applications. In *IEEE 23rd International Symposium on Consumer Technologies, ISCT 2019, Ancona, Italy, June 19-21, 2019*, pages 216–219. IEEE, 2019.

- [56] Emilio Sulis and Antonio Di Leva. Public health management facing disaster response: a business process simulation perspective. In *Proceedings of the 2018 Winter simulation Conference*, pages 2792–2802. Winter Simulation Conference, 2018.
- [57] Emilio Sulis, Llio Humphreys, Fabiana Venero, Ilaria Angela Amantea, Davide Audrito, and Luigi Di Caro. Exploiting co-occurrence networks for classification of implicit inter-relationships in legal texts. *Information Systems*, 106:101821, 2022.
- [58] Emilio Sulis, Llio Humphreys, Fabiana Venero, Ilaria Angela Amantea, Luigi Di Caro, Davide Audrito, and Stefano Montaldo. Exploring network analysis in a corpus-based approach to legal texts: A case study. In *COUrT@ CAiSE*, pages 27–38, 2020.
- [59] X. Sun and L. Robaldo. On the complexity of input/output logic. *The Journal of Applied Logic*, 25:69–88, 2017.
- [60] Wil MP Van der Aalst. Business process management: a comprehensive survey. *ISRN Software Engineering*, 2013, 2013.
- [61] Wil MP Van der Aalst, Joyce Nakatumba, Anne Rozinat, and Nick Russell. Business process simulation. In *Handbook on BPM 1*, pages 313–338. Springer, 2010.

EXPLAINABLE REASONING WITH LEGAL BIG DATA: A LAYERED FRAMEWORK

GRIGORIS ANTONIOU

University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK
g.antoniou@hud.ac.uk

KATIE ATKINSON

University of Liverpool, UK
K.M.Atkinson@liverpool.ac.uk

GEORGE BARYANNIS

University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK
g.bargiannis@hud.ac.uk

SOTIRIS BATSAKIS

*Technical University of Crete, Chania, 73100, Greece and University of
Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK*
s.batsakis@hud.ac.uk

LUIGI DI CARO

University of Turin, Turin, Italy
dicaro@di.unito.it

GUIDO GOVERNATORI

Independent researcher
guido@governatori.net

LIVIO ROBALDO

*Legal Innovation Lab Wales, Hillary Rodham Clinton School of Law, Swansea
University, Singleton Park, Swansea, SA2 8PP, UK*
livio.robaldo@swansea.ac.uk

GIOVANNI SIRAGUSA

University of Turin, Via Pessinetto 12, 10149 Torino, Italy
siragusa@di.unito.it

ILIAS TACHMAZIDIS

University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, UK
i.tachmazidis@hud.ac.uk

Abstract

Traditionally, computational knowledge representation and reasoning focused its attention on rich domains such as the law. The main underlying assumption of traditional legal knowledge representation and reasoning is that knowledge and data are both available in main memory. However, in the era of big data, where large amounts of data are generated daily, an increasing range of scientific disciplines, as well as business and human activities, are becoming data-driven. This article summarises existing research on legal representation and reasoning in order to uncover technical challenges associated both with the integration of rules and databases and with the main concepts of the big data landscape. These challenges lead naturally to future research directions towards achieving large scale legal reasoning with rules and databases.

1 Introduction

Since the emergence of computational knowledge representation and reasoning (KR), the domain of law has been a prime focus of attention as it is a rich domain full of explicit and implicit representation phenomena. From early Prolog-based approaches [49, 51] to elaborate logic-based mechanisms for dealing with, among others, notions of defeasibility, obligation and permission, the legal domain has been an inspiration for generations of KR researchers [3, 19, 36, 52].

Knowledge representation has been used to provide formal accounts of legal provisions and regulations, while reasoning has been used to facilitate legal decision support and compliance checking. Despite the variety of approaches used, they all share a common feature: the focus has always been on capturing elaborate knowledge phenomena while the data has always been small. As a consequence, one underlying assumption has been that all knowledge and data are available in main memory. This assumption has been reasonable until recently, but can be questioned with the emergence of *big data*. We now live in an era where unprecedented amounts of data become available through organisations, sensor networks and social media. An increasing range of scientific disciplines, as well as business and human activities, are becoming data-driven.

Since legislation is at the basis of and regulates our everyday life and societies, many examples of big data such as medical records in e-Health or financial data, must comply with, and are thus highly dependent on, specific norms. For instance, a sample database related to the US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) contains over 3 million records to cover only the first quarter of 2014 [34]. Any standard reasoning system would reach its limits if data over longer periods of time need to be audited.

Another source of huge amounts of data related to law is the financial domain, in which millions of transactions take place every single day and are subject to regulation on, among others, taxation, anti money laundering, consumer rights and data protection. While data mining is being used in the financial domain, it is arguably an area that would benefit from legal reasoning directly related to relevant legislation. This might indicatively entail checking for and ensuring compliance with reporting requirements, or traversing across financial transaction databases to check for potential violations of legislations.

Similarly, building applications and property/site development are covered by a variety of local and national laws and regulations. To develop and assess relevant applications, it may be necessary to consider the legal requirements in conjunction with geodata relating to morphology of the site and its surroundings, use of space and so on.

Industries in the aforementioned and other domains are feeling increasingly overwhelmed with the expanding set of legislation and case law available in recent years, as a consequence of the global financial crisis, among others. Consider, for example, the European Union active legislation, which was estimated to be 170,000 pages long in 2005 and is expected to reach 351,000 pages by 2020 assuming that legislation trends continue at the same rate [39]. As the law becomes more complex, conflicting and ever-changing, more advanced methodologies are required for analysing, representing and reasoning on legal knowledge.

While, the term “big data” is usually associated with machine learning, we argue that particularly in law there is also a need for symbolic approaches. Legal provisions and regulations are considered as being formal and legal decision making requires clear references to them. Stated another way, in the legal domain there is also a need for *explainable artificial intelligence*, as it has always been done in legal reasoning.

So what are the implications of this big data era on legal reasoning? On the one hand, as already explained above, a combination of legal reasoning with big data opens up new opportunities to provide legal decision support and compliance checking in an enhanced set of applications. On the other hand, there are new technical challenges that need to be addressed when faced with big data:

- Rules and data integration: while big data is stored in databases of various forms, reasoning is often performed using rule engines. Integrated solutions are necessary so that rule engines can seamlessly access and reason with big data in large scale databases.
- Volume: When the amount of data is huge, one cannot assume that all data is available in main memory. Hence, any approach that relies on this assumption needs to be adapted in order to work on larger scales.
- Velocity: In applications where one wishes to perform decision making close to the time data is generated, the dynamicity of data needs to be taken into account.
- Variety: In many applications, there is a need for a uniform manner of accessing and reasoning with data from disparate, heterogeneous sources, following different formats and structures.

The aim of this article is to present the state of the art in legal reasoning with rules and databases and explore the challenges faced by existing approaches when moving to larger scales and when integrating rule-based and database systems. In doing so, the article aims to stimulate the evolution of the area of legal reasoning so that it becomes more relevant in the new data-driven era.

The remainder of this article is organised as follows. Section 2 provides an overview of previous research in legal representation and reasoning. Section 3 discusses the application of legal reasoning in practice, first dealing with case studies of increasing scale, then discussing the integration of rules and databases and a possible solution through the RuleRS system. Then, Section 4 provides a description of technical challenges arising both from the integration of rules and databases and large scale case studies. Finally, Section 5 summarises findings and briefly discusses their importance.

2 Legal Representation and Reasoning Approaches

2.1 Rule-based Approaches

A quite significant subset of legal representation and reasoning approaches relies on logic-based representation and rule-based reasoning. The benefits of rule-based approaches stem mainly from their naturalness, which facilitates comprehension of the represented knowledge [38]. Rules, representing domain knowledge, are normally in the “IF conditions THEN conclusion” form; in the legal domain, conditions are

the norms and consequence is the legal effect. To apply rule-based reasoning in the legal domain, the meaning of legal texts needs to be interpreted and modelled, in order to transform the legal norms to logical rules for permitting reasoning [16].

According to [40], the main advantages of rule-based approaches are:

- compact representation of general knowledge,
- natural knowledge representation in the form of if-then rules that reflect the problem-solving procedure explained by the domain experts,
- modularity of structure where each rule is an independent piece of knowledge
- separation of knowledge from its process,
- justification of the determinations by explaining how the system arrived at a particular conclusion and by providing audit trails.

There are, however, a number of issues that pertain to the knowledge acquisition bottleneck, or inference efficiency, especially for large scale reasoning. Sections 2.2 to 2.4 summarise the most important rule-based legal reasoning approaches.

2.2 Early Logic-based Approaches

The earliest well-established approach to rule-based legal reasoning involved the use of subsets of first-order logic for knowledge representation and Prolog-based reasoning. The most prominent example is Sergot *et al.*'s seminal work on the British Nationality Act [51], where the authors expressed legal knowledge in the form of extended Horn logic programs that allow negation as failure. The authors present an excellent account of the intricacies of encoding actual legislation as rules, especially with regard to the treatment of negation and cases where double negation is introduced.

Subsequent work [35] focused, among others, on the encoding of exceptions within a particular legislation, representing them explicitly by negative conditions in the rules. While this is suitable for self-contained and stable legislation, it may require some level of rewriting whenever previously unknown exceptions (or chains of exceptions) are introduced or discovered. Moreover, in both of these works deontic concepts such as permission or obligation which are a common occurrence in legislation, have to be represented explicitly within predicate names. This is an expected characteristic when legal knowledge representation relies on standard predicate logic [9].

2.3 Description Logic-based Approaches

Following the advent of the Semantic Web, several research efforts focused on examining whether description logics and ontologies are suitable candidates for representing and reasoning about legislation. An ontology is defined a formal, explicit specification of a shared conceptualization [54]. The reusability and sharing features of ontologies are of critical importance to the legal reasoning domain, due to the complexity involved in legal documents. This complexity can be viewed from two different perspectives [20]:

- The language used in legal document is complex, especially the problem of open texture property, incomplete definition of many legal concepts of the law [18].
- the amount of information that must be collected and processed in order for lawyers or judges to evaluate a case and litigation to proceed [58].

A prime example of legal reasoning approaches using description logics is HARNES [56] (also known as OWL Judge [57]), which shows that well established sound and decidable description logic reasoners such as Pellet can be exploited for legal reasoning, if, however, a significant compromise in terms of expressiveness is made. The most important issue is that relationships can only be expressed between concepts and not between individuals: for instance, as exemplified in [56] [56], if we have statements expressing the facts that a donor owns a copyright donation and that a donor retains some rights, there is no way to express (in pure OWL) that the donor in both cases is the same individual. This can be expressed via rules (e.g., written in SWRL); however, to retain decidability these rules must be restricted to a so-called DL-safe subset [41].

Description logics provide an alternative formalisation to classical logic but still face similar issues with regard to the treatment of negation and the encoding of deontic notions. The issues related to negation are due to the fact that both classical and description logics are monotonic: logical consequences cannot be retracted, once entailed. However, the nature of law requires legal consequences to adapt in light of new evidence; any conflicts between different regulations must be accounted for and resolved [9].

2.4 Defeasible and Deontic Logic-based Approaches

The aforementioned issues led researchers to employ non-monotonic logic for the purposes of legal reasoning. An example is the Defeasible Logic framework [6], where

rules can either behave in the classical sense (*strict*), they can be defeated by contrary evidence (*defeasible*), or they can be used only to prevent conclusions (*defeaters*). Defeasible Logic has been successfully used for legal reasoning applications [7, 22, 30, 24] and it has been proven that other formalisms used successfully for legal reasoning correspond to variants of Defeasible Logic [23].

As already mentioned, the notions of permission and obligation are inherent in legal reasoning but are not explicitly defined in any of the logic systems described so far; deontic logic was introduced to serve this purpose. As formalised in [31], permission and obligation are represented by modal operators and are connected to each other through axioms and inference rules. While there has been some philosophical criticism on deontic logic due to its admission of several paradoxes (e.g., the gentle murderer), deontic modalities have been introduced to various logics to make them more suitable for reasoning with legal norms. [50] uses a combination of deontic logic and the notions of action and agents to be able to derive all possible normative positions (e.g., right, duty, privilege) and assist in policy and contract negotiation. A similar proposal [48] uses reified I/O logic to formalise the EU General Data Protection Regulation (GDPR) in 966 if-then rules (<https://github.com/dapreco/daprecokb/tree/master/gdpr>).

Defeasible Deontic Logic [28, 26] is the result of integrating deontic notions (beliefs, intentions, obligations and permissions) to the aforementioned Defeasible Logic framework. Defeasible Deontic Logic has been successfully used for applications in legal reasoning and it has been shown that it does not suffer from problems affecting other logics used for reasoning about norms and compliance [25, 24, 34]. Thus, Defeasible Deontic Logic is a conceptually sound approach for the representation of regulations and at the same time, it offers a computationally feasible environment to reason about them [28].

2.5 Case-based Approaches

Apart from rule-based approaches, a number of different solutions have been proposed for representation and reasoning in the legal domain. These are summarised next. This section discusses case-based approaches, followed by case-rule hybrids (Section 2.6) and argumentation-based approaches (Section 2.7).

Rule-based legal reasoning approaches are more suited to legal systems that are primarily based on civil law, due to their inherent rule-based nature and the fact they focus on conflicts arising from conflicting norms and not from interpretation [11]. On the other hand, common law places precedents at the center of normative reasoning, which makes case-based approaches more applicable. Case-based representations store a large set of previous cases with their solutions in the case base (or case

library) and use them whenever a similar new case has to be dealt with. The case-based system performs inference in four phases known as the CBR cycle [2]: retrieve, reuse, revise and retain. Quite often, the solution contained in the retrieved case(s) is adapted to meet the requirements of the new case.

An important advantage of case-based representation is its ability to express specialized knowledge. This allows them to circumvent interpretation problems suffered by rules (due to their generality). Also, knowledge acquisition may be slightly easier than rule-based approaches, due to the availability of cases in most application domains. However, case-based approaches face a number of issues such as the inability to express general knowledge, poor explanations and inference inefficiency, especially for larger case bases [45].

The most prominent examples of case-based legal reasoning are HYPO [8], CATO [4] and GREBE [13]. HYPO represents cases in the form of dimensions which determine the degree of commonality between two precedent cases: a precedent is more “on-point”, if it shares more dimensions with the case at hand than another. CATO replaces dimensions with boolean factors organised in a hierarchy. GREBE is actually a rule/case hybrid, since reasoning relies on any combination of rules modeling legislation and cases represented using semantic networks (a precursor to ontologies in the Semantic Web). As noted in [10], using dimensions or factors to determine legal consequences is relatively tractable, but the initial step of extracting these dimensions or factors from case facts is deeply problematic.

2.6 Hybrid Approaches

A number of attempts have been made to integrate rule-based and case-based representations [45]. Since rules represent general knowledge of the domain, whereas cases encompass specific knowledge gained from experience, the combination of both approaches turns out to be natural and useful.

In legal reasoning, such hybrid solutions are capable of addressing issues arising due to the existence of “open-textured” (i.e., not well defined and imprecise) rule terms or unstated prerequisite conditions and exceptions or circularities in rule definitions [47]. Examples of hybrid legal representation and reasoning systems are CABARET [47], DANIEL [15], GREBE [14, 12], and SHYSTER-MYCIN [1].

2.7 Argumentation-based Approaches

Regardless of the legal system applied, legal reasoning at its core is a process of

argumentation, with opposing sides attempting to justify their own interpretation. As succinctly stated in [44], legal reasoning goes beyond the literal meaning of rules and involves appeals to precedent, principle, policy and purpose, as well as the construction of and attack on arguments. This became especially apparent when Dung’s influential work on argumentation frameworks [17] started being applied in AI and law research. AI and law research has addressed this with models that are based on Dung’s influential work on argumentation frameworks. A notable example is Carneades [21], a model and a system for constructing and evaluating arguments that has been applied in a legal context. Using Carneades, one can apply pre-specified argument schemes that rely on established proof standards such as “clear and convincing evidence” or “beyond reasonable doubt”.

ASPIC+ [43] takes a more generic approach, providing a means of producing argumentation frameworks tailored to different needs in terms of the structure of arguments, the nature of attacks and the use of preferences. However, neither Carneades nor any ASPIC+ framework can be used as-is for legal reasoning: they need to be instantiated using a logic language. For instance, versions of Carneades have used Constraint Handling Rules to represent argumentation schemes, while any ASPIC+ framework can be instantiated using a language that can model strict and defeasible rules, such as those in the previously mentioned Defeasible Logic framework.

3 Legal Reasoning with Rules and Databases in Practice

As detailed in the previous section, researchers have proposed a multitude of different approaches to legal representation and reasoning, each with their own advantages and disadvantages. Focusing on rule-based approaches specifically, regardless of their individual characteristics, two major issues have not yet been adequately addressed, to the best of our knowledge. These involve handling significantly large datasets and achieving efficient integration between legal rules and databases. In this section, we explore how current rule-based legal reasoning approaches fare in relation to these issues.

3.1 Exploring Case Studies of Different Scale

As part of the MIREL project, practical legal reasoning applications were explored to complement theoretical analysis. For instance, in [9], several legal reasoning approaches were applied on real-world use cases. The approaches examined included answer set programming (ASP), defeasible logic and ASPIC+-based argumentation.

The use cases involved the presumption of innocence axioms, blockchain-based contracts use case and the FDA Adverse Event Reporting System.

The first use case (presumption of innocence) involves only a few rules but demonstrates the importance of semantics and how different formalisms deal with conflicting facts and rules, especially in the case of missing preferences between rules. The second use case is an example of rules within a contract, and is interesting due to including notions of permission, obligation and reparation. The third use case involves part of the rules applied in the FDA reporting system mentioned in the introduction. Since the number of rules and cases is big, the third use case is very relevant to the challenges of large scale reasoning.

The three formalisms were selected because of their support for complex rules involving conflicts and priorities, as is typically the case of legal reasoning, and the availability of stable tools for reasoning. All three formalisms were expressive enough for representing rules involved in the three use cases, but the user must be familiar with the underlying semantics, since in some cases the rules must be modified accordingly in order to achieve the desired behaviour. But besides their differences, the three approaches can form the basis of a large scale reasoning implementation.

The advantage of ASP is its expressiveness since it offers support for disjunction, strong negation and negation as failure and additional constructs such as aggregation functions; however, ASP reasoning has high computational complexity. Argumentation and defeasible logic offer reasoning with lower complexity, but argumentation has significantly restricted expressiveness. Overall, defeasible logic seems to provide the best trade-off between expressiveness and complexity.

The most complex use case in [9], a subset of FDA Adverse Event Reporting System, when implemented contains approximately 100 rules for all three formalisms. Reasoning times for three formalisms did not exceed a few seconds. This means that reasoning is efficient for hundreds of rules, but challenges may arise for even larger rule sets or in case reasoning results in one rule set depend on completing reasoning on another set. The main bottleneck identified, however, is representation, since manual encoding of rules and case related facts is time consuming and requires expertise in knowledge representation, and specifically in the formalism used for reasoning.

3.2 Integration Between Legal Rules and Databases

For many applications, necessary data is stored in (relational) databases. Various organizations may use the data from existing databases to comply with various regulations and guidelines, take decisions and create reports based on regulations (and other normative and legislative documents). For example, Australian financial

institutions are subject to Financial Sector (Collection of Data) Act 2001, with regard to what (financial) information to report to the relevant regulators (e.g., Australian Prudential Regulator Authority); government departments and agencies are required to comply with the Public Governance Performance and Accountability Act 2013 and Public Governance Performance and Accountability Rule 2014 for their annual financial reporting. The requirements about what, when and in what forms to comply (and related exceptions) are given in the (relevant) regulations while the (financial and other) data is stored in the databases of the institutions that have to generate reports about the data using legal reasoning.

Accordingly, in these scenarios, one has to perform some legal reasoning (for example to understand what are the actual requirements that apply in a given case) based on the information stored in enterprise databases. In fact, legal reasoning consists of five elements which lead to a decision that can be decided as either accepted or rejected ¹. The components are: issues or cases (legal), rules, facts, analysis and conclusion. The argument for a particular issue has to align with the legal rule and relevant facts corresponding to the rule. Overall, the process is analysed and apply the facts from database to the rules for generating a conclusion. Consequently, the facts stored in the enterprise database are required to apply the rules and perform legal reasoning.

Typically, database management systems involve a relatively small number of relations or files holding a large number of records, whereas rule-based systems consist of a large number of relationships with a small number of records [46]. Additionally, relational databases essentially represent knowledge in a first-order logic formalism and query languages mostly exploit first-order logic features. However, as detailed in Section 2, first-order logic is not fully suitable to represent legal knowledge. This means that in general, we cannot use solely database queries, but we have to integrate the information stored in a database with rule systems specialised in legal reasoning.

A possible solution to integrating rules with databases would be to encode and store rules in a separate application program and then align with databases. However, in this manner, it would often be difficult to adapt the program if regulations change. Additionally, it could not be guaranteed that databases and rule-based systems are consistently amended. Another solution would be to couple databases with an expert system, but this would not solve the consistency problem since data is in one system, and the rules are in another one [53]. Stonebraker suggests that rule systems integrated into the (relational) database system could be the possible solution. In this circumstance, it is required to integrate a database to serve

¹<https://groups.csail.mit.edu/dig/TAMI/inprogress/LegalReasoning.html>

legal obligations since traditional database architecture is not capable of reporting regulatory requirements.

3.3 RuleRS: A Solution to the Integration Problem

This section demonstrates RuleRS [34], a possible solution where rules and databases are integrated. Initially, we are focusing on the mapping between the two vocabularies representing rules and databases. The fundamental idea behind the mapping is that data stored in the database corresponds to facts in a legal norms and these facts can be retrieved from the database using queries (SQL, JSON). Thus, each fact corresponds to a query and a mapping is a statement that can be true or false depending on the value of its arguments/variables.

The *RuleRS* design architecture, shown in Figure 1, consists of five main system components. In particular, the key system components of RuleRS are: 1) I/O Interface, 2) Database facts 3) Formal Rules, 4) Predicates, and 5) Rule engine (SPINdle Reasoner). The following subsections provide a short outline of the RuleRS internal components and their functions.

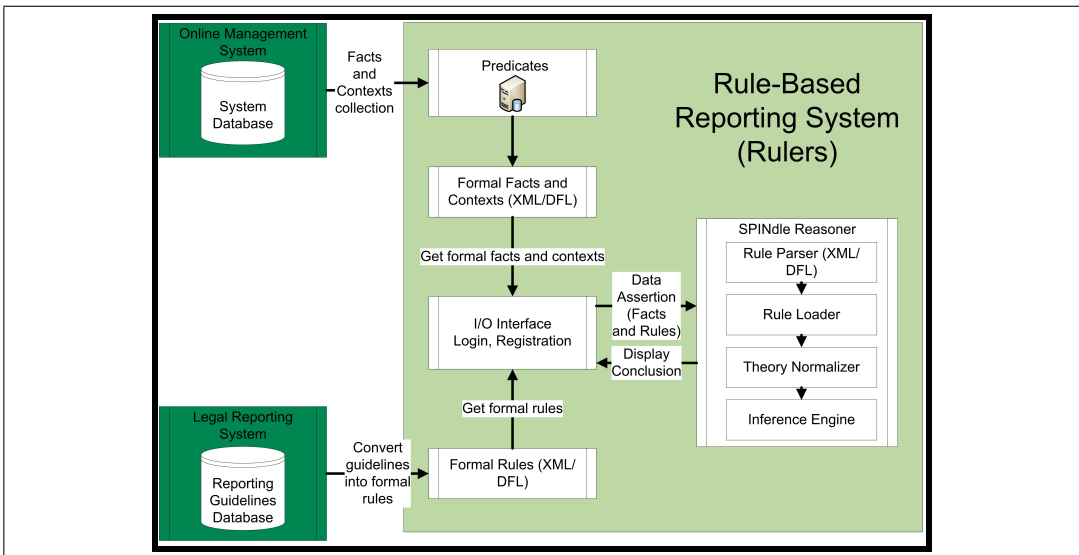


Figure 1: Rule-based Reporting System (RuleRS)

3.4 I/O Interface

The I/O Interface is implemented in Java to bridge RuleRS components and interacting with each other. The I/O interface is used to query data predicates (SQL or JSON files) and to generate facts and contexts in formal notation in Defeasible Logic syntax, and the rule engine (SPINdle reasoner) receives this as a parameter. The I/O interface also displays the final remarks or comments for each of the incidents and predicates.

3.5 Database facts

This section describes how to obtain facts from databases. In RuleRS, facts can be true or false for specific information from the database which is mapped with the literals rules. We have used either SQL or JSON (JavaScript Object Notation) syntax² syntax (or a combination of them) to represent database facts. Each of the facts are generated by queried database and send it to reasoner for further processing.

3.6 Formal Rules base

One of the prominent features of the RuleRS system is its ability to perform reasoning based on legal requirements. As we alluded to in the introduction, such regulatory requirements are represented as formal rules in Defeasible Deontic Logic [28, 26]. To enable their use with the rule engine used by RuleRS (SPINdle, see the next section) the rules are stored in the DFL format [37]. At this stage, the rules are created manually and (semi-)automatically by legal knowledge engineers and stored in a knowledge base.

3.7 Predicates

As specified earlier, since there is no direct correspondence between the literals encoding rules and the table/attributes of the database schema, we have to establish a mapping among them to enable the integration of rules and instances in the database. We named this mapping “predicates”. The fundamental idea behind predicates is that data stored in the database correspond to facts in a defeasible theory and these facts can be retrieved from the database using queries. Thus, each fact corresponds to a SQL/JSON query and a predicate is a statement that can be true or false depending on the value of its arguments/variables. A predicate with n arguments is an $n - ary$ relation mapping literals and a set of attributes. A predicate in RuleRS corresponds to a database view, i.e.; a named query, where the name is literal to

²<http://json.org>

be used by the defeasible rules. The details are the query to be run to determine if the predicate is true or false for a given set of parameters. In case the output of the query is not empty, the predicate is true and is passed to the defeasible theory as fact.

In RuleRS, predicate consists of two components: (1) predicate name and (2) predicate details. *Predicate name* represents the action(s), condition(s) or indisputable statement(s), and passed on to the rule engine, SPINdle as defeasible fact (literal and modal literal) [27, 28, 29] or actions that have been performed. For example, the fact “There is a risk for an incident” is represented by “*riskForIncident*” and passed as “>> *riskForIncident*” to SPINdle if it is returned as true from the relational database. “Predicate details” includes the “incident details” and may be stored as an SQL statement or converted to JSON to create a bridge between the data stored in the database and the terms passed as predicates (input case) to the rule engine. The SQL or JSON statements can be created in the initialisation of RuleRS with all of the incidents along with all of the predicates for each of the incidents or dynamically add it later.

Incident ID and relevant details of the incidents are also included for each of the predicates and named the predicates with relevant incident information such as “riskForIncident.sql” (for SQL statement) or “riskForIncident.json” (for JSON Statement). The following snippet illustrates the SQL syntax adopted by RuleRS for the example of the “riskForIncident” predicate:

```
SELECT incidentID, IncidentDetails, IncidentDetails1,
IncidentDetails2 FROM tblIncident
WHERE incidentID='XXXXXX'
```

In this example, `IncidentDetails`, `IncidentDetails1`, `IncidentDetails2` are substituted for the place- holders in the “riskForIncident” predicate from relational databases for the `incidentID` `'XXXXXX'`. Using JSON, the syntax for the “riskForIncident” predicate is:

```
{"riskForIncident":
{ "incidentID": "XXXXXX",
  "IncidentDetails": "ABC",
  "IncidentDetails1": null,
  "IncidentDetails2": "XYZ"}}
```

In the next step, the records and incidents for which there is a match in the relational database are transformed into predicates to be used by the SPINdle rule engine [37], and forwarded to SPINdle for further processing using the I/O interface to make the process dynamic.

3.8 The rule reasoner

RuleRS uses SPINdle Reasoner ³ [37], a Java-based implementation of Defeasible Logic that computes the extension of a defeasible theory. SPINdle supports Modal Defeasible Logic and all types of Defeasible Logic rule, such as facts, strict rules, defeasible rules, defeaters, and superiority. In summary, SPINdle is a powerful tool which accepts rules, facts, monotonic and non-monotonic (modal) rules for reasoning with inconsistent and incomplete information. In RuleRS, SPINdle Reasoner receives the formal facts, contexts as predicates from predicate file generated for data stored in the associated relational databases and computes definite or defeasible inferences which are then displayed by the I/O interface.

4 Challenges and Future Research Directions

A number of different challenges arise when attempting to move towards large scale legal reasoning with rules and databases. Some of these challenges are directly related to the integration between rules and data and are discussed in Section 4.1. Others are linked to issues raised by large scale data and are discussed in Section 4.2.

4.1 Integration between Rules and Databases

In [53], three possible forms that bring rules with database systems are discussed:

- rule policy can be written down in a booklet and distributed to people,
- the rules can reside in an application program which accesses the databases,
- a knowledge base can reside inside the DBMS by which we can guarantee that the data is consistent with the rules

The author expected that the last form will be the one to be adopted as a major approach. However, we argue that the last form may work well for a single database with small amount of rules but poses some significant challenges for large scale legal reasoning. A number of challenges are raised when attempting to integrate rules and databases, especially at larger scales and these are detailed next.

³SPINdle Reasoner is available to download freely from <http://spin.nicta.org.au/spindle/tools.html> under LGPL license agreement (<https://opensource.org/licenses/lgpl-license>)

4.1.1 Common languages

The values encoding regulation and guidelines (legal documents) and the databases (schemas) used in conjunction with the rules are in general developed independently and are likely to have a different vocabulary in general. This may lead to “Tower of Babel” issues, due to the absence of “common languages” between regulations and databases. There is no direct correspondence between the literals used by the rules and the table/attributes of the database schema. Accordingly, we have to establish a mapping between them to enable the integration of rules and instances in the database.

4.1.2 Integrating varieties of data sources with rule engines

Another challenge involves the integration of data coming from disparate sources with rule engines. Each source could publish data in their own format and all of these formats would need to be brought together to construct schema-based conditions for rules. This is quite a big asking for knowledge engineering. Furthermore, when database schemas or rules change, schema-based indices will also be affected due to the strong coupling.

4.1.3 Inference efficiency

In the case of defeasible deontic logic in legal domain, each condition in a rule could be represented by a complex query that involves multiple selections, projections and joins across multiple tables and databases. Existing schema-based index approaches cannot address this complexity well. Furthermore, rules in the legal domain have not only dependent relationship but also defeater relationship. Together with issues such as reparation chain handling, they bring more dynamics during reasoning process which places even heavier burden to inference engines.

4.1.4 Reactive inference

The existing reasoning process in systems such as RuleRS is that the inference engine looks for rules which match facts stored in the working memory or provided by users. One rule is selected from the “conflict set” and executed to generate a new fact. Then the inference engine will continue the reasoning based on the new fact together with the previous given facts. We call this as reactive inference because the inference engine only reasons based on what is given but does not interact with databases to seek “unknown” facts proactively. Proactive inference is critically important when it is highly unlikely for users to know all facts beforehand. Furthermore, the

assumption of storing facts “in memory” does not hold for large scale reasoning, as detailed in Section 4.2.2.

4.1.5 Rules as data

Rules could be treated as data and stored in database systems, to make it easier for the rules to be triggered and executed as and when required [42]. The main issue with storing rules in the database is that the database is not capable of handling deontic concepts. To correctly model the provision corresponding to prescriptive norms, we have to supplement the language with deontic operators, and the databases are not capable of handling these specific features.

Rules treated as data could create further challenges. Legal reasoning integrating rules and databases are not limited to any particular regulations. Hence, the database could be aligned to one-to-many regulations, establishing n-ary relations among these. If such rules are treated as data and stored in databases, then the task of amending them if necessary becomes even harder, since each of the rules could connect with another rule leading to nested and correlated queries. Such queries are usually avoided due to their complexity.⁴ Query maintainability and filtering also create further challenges.

4.2 Large-Scale Legal Reasoning

4.2.1 Representation

As discussed in Sections 2 and 3.1, there are several formalisms that can be used for representing legal norms and facts about cases, such as answer set programming, argumentation and defeasible logic. Although such formalisms are expressive enough for representing legal rules and efficient reasoning mechanisms and tools exist for them, encoding the rules is a complex process. For example in [9] some implementations required approximately 100 rules, and creating these rules was a time consuming process requiring expertise in logic programming. In case of large scale reasoning the encoding process will face severe scalability issues and it is a potential bottleneck for efficient large scale reasoning. Automating this process with the help of efficient natural language processing tools is an open research problem.

4.2.2 Volume

Traditional legal reasoning has been focused on storing and processing data in main memory over a single processor. This approach is indeed applicable to small legal

⁴<http://www.sqlservice.se/sql-server-performance-death-by-correlated-subqueries/>

documents. However, there is a limit on how many records an in-memory system can hold. In addition, utilising a single processor can lead to excessive processing time.

RuleRS [34] indicates that data can be processed record by record, namely querying the database and performing reasoning for each record separately. Experimental evaluation shows that this approach can evaluate each record within seconds. However, for 3 millions of records this approach requires an estimated time of 8 hours. A record by record processing approach cannot be guaranteed for any given application. Thus, in other applications where all records need to be loaded and processed together, main memory would be a hard constraint considering applicability.

Recent advances in mass parallelisation could potentially the limitations related to memory and processing time. It has been shown in literature [5] that mass parallelisation can be applied to various types of reasoning. Both supercomputers (e.g., a single large machine with hundreds of processors and a large shared main memory) and distributed settings (e.g., a large number of combined commodity machines that collectively provide multiple processors and a large main memory) can be used in order to speed up data processing. The advantages are twofold, since mass parallelisation: (a) could significantly reduce processing time as multiple cores can be used simultaneously, and (b) virtually alleviates the restriction on main memory as more memory can be easily added to the system.

4.2.3 Velocity

Financial transactions could potentially require real-time monitoring of day-to-day activity. Such functionality would depend on processing large amounts of transactions within seconds. For cases where reasoning needs to take place during a short window of time, close to the time that events take place, batch reasoning is no longer a viable solution. A prominent challenge in this situation is the efficient combination of streaming data with existing legal knowledge (e.g., applicable laws and past cases), essentially updating the latter. Stream reasoning has been studied in literature [32, 55], showing that only relatively simple rules could allow high throughput. In general, stream processing is intended for use cases where data is processed towards a single direction. However, in stream reasoning, recursive rules (i.e., rules that lead to inference loops) may lead to performance bottlenecks. In addition, within such a dynamic environment, incoming data could potentially invalidate previously asserted knowledge leading to a new set of knowledge, which would in turn change the set of conclusions.

4.2.4 Variety

One of the main challenges in large-scale legal reasoning could be the integration of data coming from disparate sources. Each source could publish data in any possible format, ranging from images of scanned pages to machine processable files. Thus, the first challenge is to translate all available data into machine processable data that can be readily stored and retrieved. Once this data transformation is achieved managing data that are stored in different formats (e.g., plain text, JSON, XML, RDF) would complicate legal reasoning as all data would need to be translated into a single format in order to have a uniform set of facts. Thus, in order to tackle data variety, all available data would need to be stored in a uniform format that would allow automated translation into facts of the chosen legal reasoning framework.

Existing work on semantic technologies can be used to address these challenges. Through the use of upper ontologies that provide definitions for a wide range of concepts, specialised legal ontologies such as LKIF [33] or bespoke ontologies, it can be ensured that all available data sources related to a large scale legal reasoning effort are eventually mapped into a unified body of knowledge.

5 Conclusion

This article argued that there is scope for research in AI and law with regard to performing effective legal reasoning when the associated knowledge and data is on a large scale and there is also a need for integration between rules and databases. A number of potential scenarios were discussed where this kind of reasoning would be useful, with use cases ranging from the pharmaceutical and financial to property development sectors.

Through a summary of state of the art and an analysis of applying rule-based legal reasoning and integrating rules and databases in practice, it becomes evident that current approaches are not fully equipped to handle large scale legal reasoning with rules and databases and face several challenges.

With regard to the problem of integration between rules and databases, the identified challenges relate to: (a) common languages; (b) integrating rule engines with various data sources; (c) inference efficiency; (d) reactive inference; and (e) rules as data. Additional challenges are encountered when moving towards larger scales, dealing with: (a) representation; (b) volume; (c) velocity; and (d) variety.

It is envisioned that these challenges, among others, will drive research on legal representation and reasoning in the near future, providing researchers at the confluence of AI and law with a multitude of potential avenues of investigation. By addressing some of these challenges, efficient, effective and successful large scale legal

reasoning with rules and databases will be achievable in the era of big data.

References

- [1] Thomas A O’Callaghan, James Popple, and Eric McCreath. Shyster-mycin: A hybrid legal expert system. In *Proceedings of the Ninth International Conference on Artificial Intelligence and Law (ICAIL-03)*, pages 103–4, 06 2003.
- [2] Agnar Aamodt and Enric Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications. IOS Press*, 7:1:39–59, 1994.
- [3] Marco Alberti, Federico Chesani, Marco Gavanelli, Evelina Lamma, Paola Mello, and Paolo Torroni. Verifiable Agent Interaction in Abductive Logic Programming: The SCIFF Framework. *ACM Trans. Comput. Logic*, 9(4):29:1–29:43, 2008.
- [4] Vincent Aleven. Using background knowledge in case-based legal reasoning: A computational model and an intelligent learning environment. *Artif. Intell.*, 150(1-2):183–237, 2003.
- [5] Grigoris Antoniou, Sotiris Batsakis, Raghava Mutharaju, Jeff Z. Pan, Guilin Qi, Ilias Tachmazidis, Jacopo Urbani, and Zhangquan Zhou. A survey of large-scale reasoning on the web of data. *Knowledge Eng. Review*, 33:e21, 2018.
- [6] Grigoris Antoniou, David Billington, Guido Governatori, and Michael J. Maher. A Flexible Framework for Defeasible Logics. In Henry A. Kautz and Bruce W. Porter, editors, *AAAI/IAAI*, pages 405–410. AAAI Press / The MIT Press, 2000.
- [7] Grigoris Antoniou, David Billington, Guido Governatori, and Michael J Maher. Representation results for defeasible logic. *ACM Transactions on Computational Logic*, 2(2):255–287, 2001.
- [8] Kevin D. Ashley. *Modeling Legal Argument: Reasoning With Cases and Hypotheticals*. The Bradford Books, MIT Press, 1990.
- [9] Sotiris Batsakis, George Baryannis, Guido Governatori, Ilias Tachmazidis, and Grigoris Antoniou. Legal Representation and Reasoning in Practice: A Critical Comparison. In *Legal Knowledge and Information Systems - JURIX 2018: The Thirty-first Annual Conference, Groningen, The Netherlands, 12-14 December 2018.*, pages 31–40, 2018.

- [10] Trevor J. M. Bench-Capon. What Makes a System a Legal Expert? In Burkhard Schäfer, editor, *JURIX*, volume 250 of *Frontiers in Artificial Intelligence and Applications*, pages 11–20. IOS Press, 2012.
- [11] Trevor J. M. Bench-Capon and Henry Prakken. Introducing the Logic and Law Corner. *J. Log. Comput.*, 18(1):1–12, 2008.
- [12] Karl Branting. A reduction-graph model of precedent in legal analysis. *Artif. Intell.*, 150:59–95, 2003.
- [13] L. Karl Branting. *Reasoning with Rules and Precedents: A Computational Model of Legal Analysis*. Springer Netherlands, 2000.
- [14] L.Karl Branting. Building explanations from rules and structured cases. *International Journal of Man-Machine Studies*, 34(6):797 – 837, 1991. AI and Legal Reasoning. Part 1.
- [15] Stefanie Brüninghaus. Daniel: Integrating case-based and rule-based reasoning in law. In *AAAI*, 1994.
- [16] Luca Cervone, Monica Palmirani, and Tommaso Ognibene. Legal rules, text and ontologies over time. In *Proceedings of the RuleML2012@ECAI Challenge, at the 6th International Symposium on Rules*, volume 874, 01 2007.
- [17] P. M. Dung. On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and n-Person Games. *Artificial Intelligence*, 77(2):321–357, 1995.
- [18] Anne von der Lieth Gardner. *An Artificial Intelligence Approach to Legal Reasoning*. MIT Press, Cambridge, MA, USA, 1987.
- [19] Marco Gavanelli, Evelina Lamma, Fabrizio Riguzzi, Elena Bellodi, Riccardo Zese, and Giuseppe Cota. Abductive logic programming for normative reasoning and ontologies. In *JSAI-isAI Workshops*, volume 10091 of *Lecture Notes in Computer Science*, pages 187–203, 2015.
- [20] El Ghosh. *Automation of legal reasoning and decision based on ontologies*. PhD thesis, INSA de Rouen, 2018.
- [21] T. F. Gordon, H. Prakken, and D. N. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10-15):875–896, 2007.

- [22] Guido Governatori. Representing business contracts in RuleML. *International Journal of Cooperative Information Systems*, 14(2-3):181–216, June-September 2005.
- [23] Guido Governatori. On the relationship between Carneades and defeasible logic. In *Proceedings of the ICAIL 2011*, pages 31–40. ACM, 2011.
- [24] Guido Governatori. The Regorous approach to process compliance. In *2015 IEEE 19th International Enterprise Distributed Object Computing Workshop*, pages 33–40. IEEE Press, 2015.
- [25] Guido Governatori and Mustafa Hashmi. No time for compliance. In *Enterprise Distributed Object Computing Conference (EDOC), 2015 IEEE 19th International*, pages 9–18. IEEE, 2015.
- [26] Guido Governatori, Francesco Olivieri, Antonino Rotolo, and Simone Scanapiego. Computing Strong and Weak Permissions in Defeasible Logic. *J. Philosophical Logic*, 42(6):799–829, 2013.
- [27] Guido Governatori and Antonino Rotolo. Defeasible logic: Agency, intention and obligation. In *Proceedings of the DEON 2004*, number 3065 in LNCS, pages 114–128. Springer, 2004.
- [28] Guido Governatori and Antonino Rotolo. Bio logical agents: Norms, beliefs, intentions in defeasible logic. *Autonomous Agents and Multi-Agent Systems*, 17(1):36–69, 2008.
- [29] Guido Governatori and Antonino Rotolo. A conceptually rich model of business process compliance. In *Proceedings of the APCCM 2010*, number 110 in CRPIT, pages 3–12. ACS, 2010.
- [30] Guido Governatori and Sidney Shek. Regorous: A business process compliance checker. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, pages 245–246, 2013.
- [31] Risto Hilpinen. Deontic logic. In Lou Goble, editor, *The Blackwell Guide to Philosophical Logic*. Wiley-Blackwell, 2001.
- [32] Jesper Hoeksema and Spyros Kotoulas. High-performance Distributed Stream Reasoning using S4. In *Proceedings of the 1st International Workshop on Ordering and Reasoning*, 2011.

- [33] Rinke Hoekstra, Joost Breuker, Marcello Di Bello, Alexander Boer, et al. The LKIF Core Ontology of Basic Legal Concepts. *LOAIT*, 321:43–63, 2007.
- [34] Mohammad Badiul Islam and Guido Governatori. RuleRS: a rule-based architecture for decision support systems. *Artificial Intelligence and Law*, 2018.
- [35] Robert Kowalski and Anthony Burton. WUENIC - A Case Study in Rule-Based Knowledge Representation and Reasoning. In Manabu Okumura, Daisuke Bekki, and Ken Satoh, editors, *JSAI-isAI Workshops*, volume 7258 of *Lecture Notes in Computer Science*, pages 112–125. Springer, 2011.
- [36] Brian Lam and Guido Governatori. Towards a model of UAVs navigation in urban canyon through defeasible logic. *Journal of Logic and Computation (JLC)*, 23(2):373–395, 2013.
- [37] Ho-Pun Lam and Guido Governatori. The Making of SPINdle. In Guido Governatori, John Hall, and Adrian Paschke, editors, *RuleML*, volume 5858 of *Lecture Notes in Computer Science*, pages 315–322. Springer, 2009.
- [38] Antoni Ligeza. *Logical Foundations for Rule-Based Systems, 2nd Ed.* Springer, 01 2006.
- [39] Vaughne Miller. How much legislation comes from Europe? House of Commons Library Research Paper, 10-62, 13 October 2010.
- [40] M. Negnevitsky. *Artificial Intelligence: A Guide to Intelligent Systems.* Addison-Wesley, 2002.
- [41] Bijan Parsia, Evren Sirin, Bernardo Cuenca Grau, Edna Ruckhaus, and Daniel Hewlett. Cautiously Approaching SWRL. Preprint submitted to Elsevier Science, 2005.
- [42] V. Paul and R. V. Polamraju. Rule management and inferencing in relational databases. In *IEEE Proceedings of the SOUTHEASTCON '91*, pages 695–697 vol.2, April 1991.
- [43] Henry Prakken. An Abstract Framework for Argumentation with Structured Arguments. *Argument and Computation*, 1(2):93–124, 2009.
- [44] Henry Prakken and Giovanni Sartor. Law and logic: A review from an argumentation perspective. *Artif. Intell.*, 227:214–245, 2015.
- [45] Jim Prentzas and Ioannis Hatzilygeroudis. Categorizing approaches combining rule-based and case-based reasoning. *Expert Systems*, 24:97–122, 2007.

- [46] Tore Risch, René Reboh, Peter E. Hart, and Richard O. Duda. A functional approach to integrating database and expert systems. *Commun. ACM*, 31(12):1424–1437, 1988.
- [47] Edwina L. Rissland and David B. Skalak. Cabaret: Rule interpretation in a hybrid architecture. *International Journal of Man-Machine Studies*, 34:839–887, 1991.
- [48] Livio Robaldo and Xin Sun. Reified Input/Output logic: Combining Input/Output logic and Reification to represent norms coming from existing legislation. *Journal of Logic and Computation*, 27(8):2471–2503, 04 2017.
- [49] Ken Satoh, Kento Asai, Takamune Kogawa, Masahiro Kubota, Megumi Nakamura, Yoshiaki Nishigai, Kei Shirakawa, and Chiaki Takano. PROLEG: An Implementation of the Presupposed Ultimate Fact Theory of Japanese Civil Code by PROLOG Technology. In *JSAI-isAI Workshops*, volume 6797 of *Lecture Notes in Computer Science*, pages 153–164. Springer, 2010.
- [50] Marek J. Sergot. A computational theory of normative positions. *ACM Trans. Comput. Log.*, 2(4):581–622, 2001.
- [51] Marek J. Sergot, Fariba Sadri, Robert A. Kowalski, F. Kriwaczek, Peter Hammond, and H. T. Cory. The British Nationality Act as a Logic Program. *Commun. ACM*, 29(5):370–386, 1986.
- [52] Mark Snaith and Chris Reed. TOAST: Online ASPIC+ implementation. In Bart Verheij, Stefan Szeider, and Stefan Woltran, editors, *Proc. of the 4th International Conference on Computational Models of Argument (COMMA 2012)*, volume 245 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2012.
- [53] Michael Stonebraker. The integration of rule systems and database systems. *IEEE Trans. Knowl. Data Eng.*, 4:415–423, 1992. This paper provides a survey on rule and database integration for a decade. The author discuss possible issue with separating two systems as consistancy of the data and knowledge base is not guranteed. Instead of coupling two system it is better to intergrate two systems. The paper also discuss the classification and implementation of DBMS rules system.
- [54] Rudi Studer, V.Richard Benjamins, and Dieter Fensel. Knowledge engineering: Principles and methods. *Data and Knowledge Engineering*, 25(1):161 – 197, 1998.

- [55] Jacopo Urbani, Alessandro Margara, Cerial J. H. Jacobs, Frank van Harmelen, and Henri E. Bal. DynamiTE: Parallel Materialization of Dynamic RDF Data. In Harith Alani, Lalana Kagal, Achille Fokoue, Paul T. Groth, Chris Bie-mann, Josiane Xavier Parreira, Lora Aroyo, Natasha F. Noy, Chris Welty, and Krzysztof Janowicz, editors, *The Semantic Web – ISWC 2013*, volume 8218 of *Lecture Notes in Computer Science*, pages 657–672. Springer, 2013.
- [56] Saskia Van de Ven, Joost Breuker, Rinke Hoekstra, and Lars Wortel. Automated Legal Assessment in OWL 2. In Enrico Francesconi, Giovanni Sartor, and Daniela Tiscornia, editors, *JURIX*, volume 189 of *Frontiers in Artificial Intelligence and Applications*, pages 170–175. IOS Press, 2008.
- [57] Saskia Van de Ven, Rinke Hoekstra, Joost Breuker, Lars Wortel, and Abdallah El-Ali. Judging Amy: Automated Legal Assessment using OWL 2. In Catherine Dolbear, Alan Ruttenberg, and Ulrike Sattler, editors, *OWLED*, volume 432 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.
- [58] Michelle J. White. Legal complexity and lawyers’ benefit from litigation. *International Review of Law and Economics*, 12(3):381 – 395, 1992.

ARTIFICIAL INTELLIGENCE AND SPACE LAW

GEORGE ANTHONY LONG
Legal Parallax, LLC, USA
gal@spacejurist.com

CRISTIANA SANTOS
Utrecht University, The Netherlands
c.teixeirasantos@uu.nl

LUCIEN RAPP
Université Toulouse Capitole 1, SIRIUS Chair, France
lucien.rapp@ut-capitole.fr

RÉKA MARKOVICH
University of Luxembourg, Luxembourg
reka.markovich@uni.lu

LEENDERT VAN DER TORRE
University of Luxembourg, Luxembourg
leon.vandertorre@uni.lu

Abstract

In the next few years, space activities are expected to undergo a radical transformation with the emergence of new satellite systems and new services incorporating artificial intelligence and machine learning. This transformation covers a wide range of innovations from autonomous objects with their own decision-making power to increasingly sophisticated services exploiting very large volumes of information from space. This article identifies some of the legal and ethical challenges linked to their use. These legal and ethical challenges call for solutions that the international treaties currently in force are not able to determine and implement sufficiently. For this reason, a methodology must be developed that makes it possible to link intelligent systems and services to a system of applicable rules. Our proposed methodology refers to existing legal AI-based tools amenable to making space law actionable, interoperable and machine readable for future compliance tools.

The work conducted by Cristiana Santos was carried out under a service contract with the University of Luxembourg prior to her joining Utrecht University.

1 Introduction

Governance of space activities is faced with progressive transformation associated with the emergence of satellite systems and space-based services employing artificial intelligence (AI), including machine learning (ML). This article identifies and examines some fundamental legal challenges related to the use of AI in the space domain. Ascertaining such legal challenges requires ascertaining that space systems and services that use AI are linked to a system of governing rules and guiding legal principles.

The nature of the space and satellite industry presents a quintessential use case for AI. Virtually all space activities and ventures constitute fertile ground for employing AI. In fact, AI is ready for use in Earth orbit activities like active debris removal (ADR), near Earth ventures like abiotic resource extraction, and deep space exploration. Generally, AI applications take place in two principal ways:

- *autonomous robots* (or space objects). Whether in the form of autonomous spacecraft or satellite constellations, autonomous or intelligent space objects have the ability to not only collect, analyse, and use data for information and operational purposes but also to go where no human has ever gone or could go to collect probes and data. This application also includes autonomous spacecraft and swarm intelligence that assist in space activities such as: mining and using abiotic resources, exploring, in-orbit servicing (IoS), active debris removal, and protecting space assets — which includes protecting themselves from rogue and unknown natural space objects;
- analysing and, if necessary, acting upon *big data from space* related to: (1) debris monitoring, (2) self-preservation against potential threats from rogue and unknown natural objects in the space domain and perceived threats from other human-manufactured objects, (3) predictive analytics using very high resolution (VHR) satellite imagery, (4) real-time geospatial data analysis, and (5) analysis of data products derived from the convergence of a wide spectrum of sources (e.g. satellites, drones, the Internet of Things (IoT), unmanned aerial vehicles (UAV) imagery and UAV location data). Big data from space also enables the provision of space cloud computing services where data is stored on space-based assets.¹ Indeed, the development of AI-based technologies combined with space data can enhance the production, storage, access and dissemination of data in outer space and on Earth.

¹This is done to increase data capacity, reduce the cost of services and allow real-time access to data storage.

Space is undergoing seismic shifts driven by: New Space (promoting a Smart, Fast and Now Space) [31]; the Google, Amazon, Facebook and Apple (GAFA) web giants; venture capital firms; and start-ups. There has been significant growth in the number of space activities, space objects, and space actors. However, **new challenges** are emerging in the course of such active exploration and use while deploying AI in space. Harnessing (using and misusing) AI (and specifically ML) technologies to access and explore outer space, and engaging in space-enabled downstream commercial applications and services will, in all likelihood, **lead to a wide range of intended and unintended consequences** that cannot be downplayed or disregarded. The following risks merit attention:

(i) privacy issues associated with the use of these technologies, e.g. citizen tracking and surveillance; potential re-identification of individuals; function creep; fake imagery; biased automatic decision-making and unjust discrimination based on nationality, gender, race and geographic localisation; lack of transparency etc.; and

(ii) liability issues emerging from the potential for damage caused by, for example, collisions with autonomous spacecraft or hacking/malware aimed at weaponising AI, and the consequences of such damage for space data (security problems for sensitive data stored in outer space, and malicious data capture).

These risks are more acute when **important facets of the space field are acknowledged**. First, space is a service- and needs-oriented market, dominated mostly by demand and competitive **industry logics**, and **without a centralised regulatory body** to govern it. Second, **space activities on Earth will have increasingly pervasive repercussions** as the benefits and solutions that space provides for the problems and needs of mankind (transport, smart city management, security, agriculture, climate change monitoring etc.) become ubiquitous.² The European Space Agency (ESA) estimates that for every euro spent on the sector, six euros benefit society. This correlation reflects the **Earth's more marked dependence on space-based services**.

The range of these space-based services – many of them AI-enabled – requires consideration of a wide range of legal and regulatory issues that cannot be answered by the space industry alone. However, **UN space treaties** leave much uncertainty as to which AI uses and activities are permitted in space. Clearly, there is a need to develop or reinterpret the rules of the road' to enable commercial and civilian actors to have continued and legally compliant access to space. The principal objectives of this article are as follows:

1. to identify and discuss potential risks and challenges associated with implementing AI and ML in space;

²According to Hon. Philip E. Coyle, Senior Advisor, Center for Defense Information [1]

2. to analyse the extent to which the current *corpus iuris spatialis* (from the 1970s) can still provide answers to these risks and challenges, and choose which methodology to follow going forward; and
3. to discuss how AI-based legal tools can support space law.

In accordance with these objectives, **Section 2** examines the specifics of AI in space, describes the distinct features of AI on Earth, and demonstrates the usefulness and benefits of AI in space. **Section 3** analyses some legal, ethical and governance risks associated with AI in space. **Section 4** discusses limitations in the current space law legal framework relating to AI in space. **Section 5** offers a methodological approach to determining the legal regime applicable to AI in space. **Section 6** discusses AI-based tools that enable knowledge representation and reasoning about space law. **Section 7** summarises our analysis of AI in space.

2 Contextual Dynamics of Space and the Specifics of AI in Space

Space technology, data and services have become indispensable to the daily lives of Europeans and most people on Earth. Space-based services and activities also play an essential role in preserving the strategic and national security interests of many States. The European Union (EU) is seeking to cement its position as one of the major spacefaring powers by allowing extensive freedom of action in the space domain to encourage scientific and technical progress and support the competitiveness and innovation capacity of its space sector industries.

To boost the EU's space leadership beyond 2020, the European Parliament and Council proposed a regulation to establish the EU's space programme and the European Union Agency for the Space Programme.³ The proposed budget allocation of **EUR 16 billion for the post-2020 EU space programme**⁴ was received by the European space industry as a clear and strong signal of the EU's political willingness to reinforce the EU's leadership, competitiveness, sustainability and autonomy in space.⁵ AI is one area where the EU is exerting its leadership in space.

³In a vote on 17 April 2019, the European Parliament endorsed a provisional agreement reached by co-legislators on the EU Space Programme for 2021-2027, bringing all existing and new space activities under the umbrella of a single programme to foster a strong and innovative space industry in Europe. See [15]

⁴These benefits represent a return on investment for the European Union of between 10 and 20 times the cost of the programme.

⁵This budget will be used first to maintain and upgrade the existing infrastructures of Galileo

The use of AI in space capitalises on the emergence of **New Space**⁷ which is creating a more complex and challenging environment physically, technologically and operationally. The current contextual dynamics of space and the **specifics of space activities amenable to AI** are discussed below.

2.1 Contextual dynamics of space

Current space activities are defined as belonging to the *Space 4.0 era*, characterised by proactiveness and open-mindedness to both technology disruption and opportunity [22], and whose trends include big data from space (e.g. data imagery) and applied predictive and geospatial analytics. In particular, this era is supported by AI-based technology, machine learning, and the Internet of Things (IoT). IoT is expected to be pervasive by 2025. Data explosion will be driven by connected “things” with sensors deployed by mega constellations of small satellites (smallsats), such as those produced by Hiber and Astrocast.

The use of these technologies is bringing about a *digital revolution*, unlocking access to space-based benefits [86]. The space industry is now moving towards leveraging full digitalisation of its *products* (high-performance spacecraft infrastructure, onboard computers, antennas and microwave products), *new processes* (increasing production speed and decreasing failure rates), and *data uptake* (the ability to access data right away) for the purpose of data distribution, as well as data analytics, processing, visualisation and value adding. All this is enabling Earth observation (EO) to become part of the larger data and digital economy.

These space-based benefits (products/processes/data uptake) increase the *repercussions of space activities on Earth*. A growing number of key economic sectors (in particular land and infrastructure monitoring, security, the digital economy, transport, telecommunications, the environment, agriculture, and energy) use satellite navigation and EO systems.

Space democratisation and privatisation reflect access to and participation in space by spacefaring nations and non-governmental entities such as privately owned juridical entities. Among space actors, the **private sector** currently accounts for 70% of space activity.⁶ This percentage will only increase with the emergence of new private actors who, thanks to frontier technologies such as AI and the data revolution [17], are seeking commercial opportunities from the exploration and ex-

and Copernicus, so that EU systems remain on top. Second, the EU will adapt to new needs, such as fighting climate change, security, and the Internet of Things.

⁶ “Nowadays, private sector augments all segments of the space domain, from ground equipment and commercial space transportation to satellite manufacturing and Earth observation services” [92].

plotation of space and its resources.

Apart from emerging new technologies such as AI, new actors are developing *new global business models* driven by demand for satellite constellations, tourism, asteroid and lunar mining, in-situ resource utilization (ISRU) [50], fifth-generation technology (5G), in-orbit servicing (IoS), three-dimensional (3D) printing of satellite parts (e.g. solar panels etc.), and commercial space stations, among others. These new business segments⁷ are leveraging the space economy. The space economy is expanding enormously, with predictions that it will generate revenues of US\$ 1.1—2.7 trillion or more by 2040 [93].

New high-end technologies and small-satellite design characterise the current landscape of the space industry. Smaller, lightweight satellites based on affordable off-the-shelf hardware, less expensive spacecraft (small-, nano- and pico-satellites) can be replaced more easily, thereby stimulating rapid improvements in technology [78]. This, as well as the fact that thousands of these satellites can be launched into mega constellations, opens up the possibility for more missions and applications using space infrastructure.

2.2 Specifics of space amenable to AI

It is still important to consider the specifics of how AI is used in outer space and why that usage is distinct from terrestrial usage:

i. Space conditions are difficult and are only amenable to AI machines. Space is a remote, hostile and hazardous environment for human life⁸, and certain activities are impossible for humans to carry out and survive the ordeal. This renders space technologies dependent on AI-based technologies and processes [85]. AI-based technologies are a good fit for operational decision-making because they are robust, resilient, adaptable and responsive to changing threats.

ii. Upstream and downstream impact of AI in space. AI in a fast-approaching future will impact all sectors of the space industry: launch, constellation control, satellite performance analysis [35], AI logic in onboard payload used in deep space applications, the downstream sector of telecommunications, and Earth observation in commercial applications such as image classification and predictive analysis of phenomena.

iii. Autonomy of intelligent space objects. Using AI, a spacecraft may

⁷And others, like scalability and agility, media/advertising, business-to-consumer (B2C), vertical integration, and position in value chains.

⁸Due to e.g. difficult accessibility, the complexity of extra-atmospheric missions, the extreme physical and climatic conditions, new gravitational forces, different temperature ranges and unknown collisions with dust or asteroids.

be able [101] to recognise a threat, capture relevant data, understand the nature of the threat, and counteract it or take evasive action. The spacecraft can even share its newly acquired knowledge with other satellites. For example, a rover exploring Mars that needs to contact Earth takes up to 24 minutes to pass a signal. That leaves rather a long time for making crucial decisions that can affect the mission, which is why engineers are increasingly providing space robots with the ability to make decisions themselves [85]. With AI, space objects can, without any human involvement, collect and analyse data and decide what information to send back to Earth and when. An AI system can predict and self-diagnose problems so that it can fix itself while continuing to perform [35]. When collisions occur between intelligent space objects and debris, this brings issues relating to liability to the fore, some of which are discussed in Sections 3.1 and 4.

iv. Asset protection. Space assets could be protected with the development of an AI-based automatic collision avoidance system that can assess the risk and likelihood of in-space collisions, improve the process of deciding whether an orbital manoeuvre is needed, and transmit warnings to other space objects that are potentially at risk [23].

v. Big Data from space. Big data from space [82] refers to massive spatio-temporal Earth and space observation data collected by a variety of sensors (ranging from ground-based to space-borne) and their synergy with data from other sources and communities. Big data from space combined with “big data analytics” delivers “value” in terms of volume, velocity, variety and veracity. Traditional tools cannot capture, store, manage and analyse huge volumes of data to the same extent. Geospatial intelligence is one of many ways artificial intelligence is used in outer space. The term refers to employing AI to extract and analyse images and other geospatial information relating to terrestrial, aerial, and/or spatial objects and events. It allows events like disasters, the migration and safety of refugees, and agricultural production to be interpreted in real time. These aspects are analysed in Section 3.2 of this article.

3 Risks of AI in Space

AI in space is leading to a gradual shift from “computer-assisted human choice and human-ratified computer choice” [16] towards non-human analysis, decision-making and actions. The emerging deployment and use of intelligent space objects⁹ brings

⁹A space object is limited to the object, including its component parts, that was “launched” into space. The issue can become a bit murkier if intelligent space objects can be manufactured and deployed in situ in outer space.

novel challenges to the current space law regime, especially when (and not if) the use of such objects for the purposes of AI systems or services causes terrestrial and/or extraterrestrial injury such as violation of privacy rights, violation of data protection requirements, or injury resulting from collision with a space object [87].

The space law treaty regime consists of the foundational Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (the “Outer Space Treaty” or OST)¹⁰ and its progeny treaties. The OST embeds the cornerstone principles of current international space law jurisprudence [95]. Its principles have been elaborated on in the following progeny treaties: the Agreement on the Rescue of Astronauts, the Return of Astronauts and the Return of Objects Launched into Outer Space (the “Rescue Agreement”)¹¹, the Convention on International Liability for Damage Caused by Space Objects (the “Liability Convention”)¹², the Convention on Registration of Objects Launched into Outer Space (the “Registration Convention”)¹³, and the Agreement Governing the Activities of States on the Moon and Other Celestial Bodies (the “Moon Treaty”)¹⁴. Liability issues associated with AI risks require analysis of the Outer Space Treaty and the Liability Convention.

3.1 Liability of intelligent space objects

Liability under the space law treaty regime is based on Article VII of the Outer Space Treaty, which is the genesis of the Liability Convention. Article VII imposes international liability only on the launching State.¹⁵ The Liability Convention establishes a restricted framework for assessing international liability which also applies only to launching States [47]. Determination of liability and allocation of fault is based on where the damage occurred. Article II of the Liability Convention imposes

¹⁰Entered into force Oct. 10, 1967, 18 UST 2410; TIAS 6347; 610 UNTS 205; 6 ILM 386 (1967).

¹¹Entered into force Dec. 3, 1968, 19 UST 7570; TIAS 6599; 672 UNTS 119; 7 ILM 151 (1968).

¹²Entered into force Sept. 1, 1972, 24 UST 2389; TIAS 7762; (961 UNTS 187; 10 ILM 965 (1971).

¹³Entered into force Sept. 15, 1976, 28 UST 695; TIAS 8480; 1023 UNTS 15; 14 ILM 43 (1975).

¹⁴Entered into force July 1, 1984, 1363 UNTS 3; 18 ILM 1434 (1979). The Moon Treaty is viewed differently to the other space treaties because it has not received the international ratification of the other space law treaties. Major spacefaring nations such as the United States, Russia and China have neither signed nor ratified the treaty.

¹⁵Article 1(c) of the Liability Convention defines the term “launching State” as the State that launches or procures the launch of the space object and the State from whose territory or facility the space object is launched. A non-governmental space actor does not have international liability under the Liability Convention for damage caused by the space object regardless of its culpability. This means that a State space actor can only have international liability if it comes within the definition of a “launching State”.

absolute or strict liability for damage caused by a space object on Earth or to an aircraft in flight. On the other hand, if a space object causes damage in outer space or to a celestial body, then liability is based on the degree of fault, as stipulated in Article III. This section applies these liability rules to intelligent space objects.

3.1.1 Some notes on liability and intelligent space objects

The concept of “damage” in the Liability Convention is neither comprehensive nor unambiguous. Article 1(a) defines “damage” as **loss of life, personal injury or other impairment of health; or loss of or damage to property of States or of persons, natural or juridical, or property of international intergovernmental organizations.**” This definition creates uncertainty about the parameters or scope of damage covered by the convention. It is unclear whether the damage is limited to physical damage caused by the space object [98] or whether it extends to non-kinetic harm, indirect damage and purely economic injury [46]. Similarly, the scope of the phrase “other impairment of health” is not yet settled. For instance, is the phrase limited to physical injury or does it extend to emotional and/or mental injury? Like all legal issues associated with the Liability Convention, the scope of a damage claim is resolved according to whether the definition of damage is given a restrictive or extended interpretation. Intelligent space objects, i.e. autonomous space objects utilising AI, present challenges for the strict and fault liability scheme imposed on launching States.

Article III of the Liability Convention reads as follows:

In the event of damage being caused elsewhere than on the surface of the Earth to a space object of one launching State or to persons or property on board such a space object by a space object of another launching State, **the latter shall be liable only if the damage is due to its fault or the fault of persons** for whom it is responsible.
(Emphasis added)

Intelligent space objects disrupt Article III’s fault-based liability scheme because the decisions, acts and omissions of an intelligent space object may be construed as not being the conduct of a person and may not always be attributable to a launching State.

3.1.2 Fault liability is predicated on human fault

Generally, we think of a person as a human being, but in the legal arena, the term “person” generally refers to an entity that is subject to legal rights and duties [84].

The law considers artificial entities like corporations, partnerships, joint ventures and trusts to be “persons” as they are subject to legal rights and duties, and the law sometimes recognises and imposes legal rights and duties on certain inanimate objects like ships, lands and goods, with the result that those inanimate objects are subject to judicial jurisdiction and therefore liable to judgments made against them [84]. However, the legal rights and duties imposed on artificial entities and inanimate objects flow from the actions or conduct of human beings.

This is not necessarily the case with intelligent **machines**. A machine can learn independently from human input and can make decisions based on what it has learnt and other available information, but those abilities do not necessarily equate to natural or legal personhood. As noted, the decisions and conduct of legal persons are ultimately decisions made by human beings. This means that the decisions are not based solely on intellect or data but are also the product of human factors such as consciousness, emotions and discretion [84]. Thus, the concept of legal personhood is ultimately premised on humanity, and AI-based decisions and conduct divorced from human oversight or control arguably lack such human factors [84]. Moreover, no law currently grants “personhood” to an intelligent space object. The lack of direct or indirect human considerations in the decision-making of an intelligent machine, together with the fact that such an object has no legal rights or duties under existing law, strongly suggest that decisions made by an intelligent space object are not made by a natural or legal person.¹⁶

Since fault liability under Article III of the Liability Convention is premised on a State or persons being at fault, a decision by an intelligent space object will, in all likelihood, not be the “fault of persons”. Accordingly, assessing fault liability under Article III for a decision made by an intelligent space object may very well depend on whether such a decision can be attributable to the launching State.

3.1.3 Fault liability in the absence of human oversight in the decision-making process

In general, a State’s liability for damage or injury is traceable to human acts or omissions. This basis for imposing liability appears to be inapplicable when damage or injury in outer space is caused by a machine’s own analysis, decision and course of action all carried out without human approval [43].

Liability premised on human acts or omission does not work when no particular human had the ability to prevent the damage, short of making the decision whether to utilise AI in a space object [43]. Certainly, it is substantively difficult to draw the line between relying on AI to supplant the judgement of a human decision-maker

¹⁶[46] Note 39, at page 7.

and allowing a machine, or a non-human, to decide on a course of action and go through with it [43]. To that extent, it seems that the fault-based liability of a launching State should not be premised solely on a decision to launch an intelligent space object, because such a sweeping basis for liability would effectively retard the development and deployment of intelligent space objects.¹⁷ Thus, the appropriate question would seem to be: what conduct is necessary to attribute fault liability to a State for damage caused by an intelligent space object when human oversight is not involved in the event causing the damage?

Resolving this dilemma presents novel and complex issues associated with standard of care, foreseeability and proximate cause, which are crucial elements in establishing fault (under Article III of the Liability Convention).¹⁸ This matter is further complicated by the distinct possibility that it may not be possible to ascertain how an intelligent space object has made a particular decision [46].

Nevertheless, untangling these nuanced legal obstacles may not be necessary to assess fault liability. Article VI of the Outer Space Treaty requires a State to assure that the space activities of its governmental and non-governmental entities comply with the Outer Space Treaty. It not only makes a State internationally responsible for its national activities in outer space, but also imposes a dual mandate of “authorization and continuing supervision” that is not limited to the launching State or the space actor’s home State [46].

Article VI of the Outer Space Treaty does not expressly burden the launching State with the obligation to authorise and supervise. Instead, it bestows powers of authorisation and continuous supervision on the **appropriate State**. Since neither Article VI of the Outer Space Treaty nor any other provision of the space law treaty regime defines the term “appropriate State”, or sets out any criteria for establishing the appropriate State(s), there are no agreed legal standards for determining what constitutes an “appropriate State”. Nevertheless, some scholars have stated that a launching State is generally always an appropriate party for the purposes of Article VI of the Outer Space Treaty [11]. This is a reasonable and accurate extrapolation since the liability scheme is predicated on launching State status.

Since fault liability is generally predicated on a breach of a standard of care¹⁹, the dual responsibility of “authorization and continuing supervision by the appropriate State party” arguably establishes a standard of care that a launching State

¹⁷See [46] Note 39, at page 7. See also [44].

¹⁸[46] Note 39, at page 8. While the decision to launch an intelligent space object may not be the basis for fault liability, as discussed *infra*, how the decision was made may serve as a vehicle for assessing fault liability.

¹⁹See [19], Note 3.

must comply with²⁰, especially in connection with an intelligent space object. This essentially means that **a launching State bears the responsibility for ensuring that appropriate authorisation and supervision is exercised in connection with an intelligent space object that it launches for a non-governmental entity, regardless of whether the object is owned or operated by the national entity.** The standard of care analysis, therefore, shifts from the specific event that caused the damage to examining whether the launching State exercised sufficient authorisation and supervision over the activities of the intelligent space object.

In analogy with the **due diligence**” standard under international law [19], a **flexible** and fluid standard is used when determining whether a launching State exercised sufficient authorisation and supervision. “Due diligence” is not an obligation to achieve a particular result; rather it is an obligation of conduct that requires a State to engage in sufficient effort to prevent harm or injury to another State, its nationals²¹ or the global commons (see [34]; [68]). Breach of this duty is not limited to State action, but also extends to the conduct of a State’s nationals.²² While there is “an overall minimal level of vigilance” associated with due diligence, “a higher degree of care may be more realistically expected” from States that have the ability and resources to provide it [34]. In any event, it would appear that a launching State’s standard of care entails assuring that there is some State authorisation and supervision over the space activities engaged in by the intelligent space object. However, with the **flexible** standard of care, it seems that the appropriate degree of human oversight required, if any, depends on the function of the intelligent space object.

This flexibility is consistent with the approach of the **European Commission** (EC) to artificial intelligence in general. In its White Paper on AI [26], the EC adopted the policy that **human oversight is a necessary component in the use of AI**, based on the reasoning that human oversight ensures that an “AI system does not undermine human autonomy or cause other adverse effects” [26, p. 21]. The White Paper further stipulates that human oversight requires “appropriate involvement by human beings”, which may vary depending on the “intended use of the AI system” and the “effect”, if any, it can have on people and legal entities. It then enumerates certain non-exhaustive kinds of human oversight including 1) human review and validation of an AI decision either before or immediately after the

²⁰See generally [11], Note 67.

²¹See Seabed Mining Advisory Opinion at ¶117 (Seabed Dispute Chamber of the International Tribunal of the Law of the Sea, Case No 17, 1 February 2011) and United States Diplomatic and Consular Staff in Tehran (U.S. v. Iran), 1980 I.C.J. 3, 61 - 67 (May 24).

²²See [34] Note 73 at page 243.

decision is made, 2) monitoring of the AI system while in operation and the ability to intervene in real time and deactivate it, and 3) imposing operational restraints to ensure that certain decisions are not made by the AI system. This EC policy presents a flexible framework for determining whether a launching State has met its standard of care in relation to a non-governmental intelligent space object that causes damage in outer space.

The flexible standard of due diligence can also be used by the launching State to negate or mitigate its liability for damage caused by an intelligent space object. The flexible standard will allow a launching State to argue that the home State of the non-governmental space actor has a greater degree of oversight responsibility than the launching State. Accordingly, it should be reasonable and sufficient for a launching State to rely on assurances that the non-national's home State exercises adequate authorisation and oversight procedures for its nationals' use of intelligent space objects. This shifts the supervisory obligation from the launching State to the home State of the non-governmental space actor. The home State's failure to properly exercise its standard of care may, depending upon the circumstances, mitigate or absorb the launching State's fault liability under Article III of the Liability Convention. This shift, however, is not automatic as the due diligence standard makes it dependent on the home State's technological prowess in the area of AI or its financial ability to contract out such expertise.

3.1.4 Intelligent space objects and absolute liability

Article II of the Liability Convention imposes strict liability on a launching State if a space object causes damage on the Earth's surface or to aircraft in flight. Article VI(1), however, allows **exoneration** from absolute liability if the damage results "either wholly or partially from gross negligence or from an act or omission done with intent to cause damage on the part of a claimant State or of natural or juridical persons it represents." This defence, however, may not be available if the damage resulted "wholly or partially" from the act or omission of an intelligent space object deployed or controlled by the claimant State or a natural or juridical person the claimant State represents.

Gross negligence", the mental element of an act or omission, is the product of human thought, which is absent in the machine's decision-making process. Even more so, Article VI of the Liability Convention may also defeat exoneration from absolute liability if the claimant State is able to show that the launching State's deployment of the intelligent space object breached its State responsibility under international law, including the United Nations Charter or the Outer Space Treaty. This counter-argument to the negation of absolute liability thrusts Article VII of

the Outer Space Treaty into consideration.

3.1.5 Intelligent space objects and liability under Article VII of the Outer Space Treaty

Article VII of the Outer Space Treaty imposes international liability on the launching State, without qualification or exception. Moreover, Article VII does not predicate fault liability on human involvement in the damage-causing event or fault being otherwise attributable to the launching State. Bestowing unqualified liability on the launching State may present an alternative way to obtain compensation for damage in space caused by an intelligent space object. Monetary compensation under Article VII of the Outer Space Treaty may well be pursued when fault cannot be assessed under Article III of the Liability Convention because the decision that caused the damage was not made by a person and the decision is not otherwise attributable to a launching State. The issue can also surface if a claimant State seeks financial compensation for an injury or harm caused by an intelligent space object that does not come within the meaning of “damage” under Article 1(a) of the Liability Convention. For instance, if an intelligent space object is used to interfere with, jam or hijack a commercial satellite transmission, then the financial injury suffered as a consequence of such conduct may not be compensable under the Liability Convention given its definition of “damage.” However, Article VII of the Outer Space Treaty may provide the basis for recovery in such circumstances.

Of course, a party seeking to pursue such a remedy under Article VII of the Outer Space Treaty may, in all likelihood, encounter the objection that since the Liability Convention is the progeny of Article VII of the Outer Space Treaty, the State is precluded from pursuing a remedy directly under Article VII of the Outer Space Treaty. Such an objection may be premised on the public international law principle that “when a general and a more specific provision both apply at the same time, preference must be given to the specific provision” [61]. It is unclear if this principle applies to the relationship between Article VII of the Outer Space Treaty and the Liability Convention.

Although the Liability Convention expressly proclaims that one of its principal purposes is to establish rules and procedures “concerning liability for damage caused by space objects”,²³ the treaty does not assert that its rules and procedures are exclusive when assessing liability through means other than the Liability Convention. Most importantly, neither the Outer Space Treaty nor the Liability Convention precludes recovery of damage under Article VII of the Outer Space Treaty. This

²³Liability Convention Preamble, 4th Paragraph. The other purpose is to ensure prompt payment “of a full and equitable measure of compensation to victims” in accordance with the Convention.

point is significant given that one of the general principles of international law is that what is not prohibited is permitted.²⁴ In other words, “in relation to a specific act, it is not necessary to demonstrate a permissive rule so long as there is no prohibition.”²⁵

Determining whether Article III of the Liability Convention precludes a State from having recourse to Article VII of the Outer Space Treaty for an injury caused by a space activity is, like most current space law issues, purely an academic exercise inasmuch as there is not much guidance from national or international courts, tribunals, or agencies on how to interpret the provisions of the space law treaty regime. Nevertheless, resolving the issue involves a binary choice as to whether the Liability Convention does or does not preclude resorting to Article VII of the Outer Space Treaty. Resolution of the issue will have a significant impact on whether the Liability Convention needs to be amended or supplemented to accommodate the deployment and use of intelligent space objects. Certainly, if relief can be obtained under Article VII of the Outer Space Treaty when a remedy is not available under the Liability Convention, then Article VII of the Outer Space Treaty should provide sufficient flexibility to address liability issues associated with intelligent space objects during this period of AI infancy.

3.2 Data protection and ethical challenges related to AI in space

Every year, commercially available satellite images are becoming **sharper** and are being taken more frequently. Commercially available cutting-edge imagery resolution software limit each pixel in a captured image to approximately 31 cm.²⁶ There is increasing demand from private commercial entities to lower the resolution restriction threshold to 10 cm [99, 39]. The significance of using AI with satellite imaging is best illustrated by the immediate interim export controls imposed by the United States in January 2020 to regulate the dissemination of AI software. AI software subject to these controls include those that can automatically scan aerial images to recognise anomalies or identify objects of interest such as vehicles, houses, and other structures.²⁷

Speculation abounds regarding satellite imagery that can discern car plates, individuals, and “manholes and mailboxes” [6]. In fact, in 2013, police in Oregon used

²⁴S.S. *Lotus*, P.C.I.J. Ser. A, No. 10 at 18 (1927).

²⁵[90] quoting *Accordance with International Law of the Unilateral Declaration of Independence in Respect of Kosovo* (Kosovo Advisory Opinion), Advisory Opinion, 2010 I.C.J. 403 (July 22) (declaration of Judge Simma at 2).

²⁶<http://worldview3.digitalglobe.com/>

²⁷85 Fed. Reg. 459 (January 6, 2020)

a Google Earth satellite image that showed marijuana growing illegally on a man’s property [10]. In 2018, Brazilian police used real-time satellite imagery to detect the spot where trees had been ripped out of the ground to illegally produce charcoal, and they arrested eight people in connection with the scheme [33]. In China, human rights activists used satellite imagery to show that many of the Uigur re-education camps in the Xinjiang province are surrounded by watchtowers and razor wire [100]. In one recent case, ML was used to create a system that could autonomously review video footage and detect patterns of activity at a particular location. This system was used to monitor a video of a parking lot and identify moving vehicles and pedestrians. The system established a baseline of normal activity from which anomalous and suspicious actions could be detected [2].

Even if such image and video resolution systems are not able to identify individuals or their features [75], they are **no longer in a sweet spot**. The broad **definition of personal data** in the General Data Protection Regulation (GDPR)²⁸ allows *all* information from EO data related to an *identified* or *identifiable* natural person (like location data) to be considered as personal data. The attribute “identified” refers to a known person, and “identifiable” refers to a person who has not yet been identified but whose identification is still possible. An individual is directly identified or identifiable with reference to *direct or unique identifiers*. These “direct and unique identifiers” cover data types that can be easily referenced and associated with an individual, including descriptors such as name, identification number, username, location data, the Subscriber Identity Module (SIM) cards of mobile phones, online identifiers etc., as described in Article 4(1) of the GDPR. An individual is *indirectly identifiable* by a combination of indirect (and therefore non-unique identifiers) that allow an individual to be singled out; these are less obvious information types which can be related to, or “linked” to an individual — for instance video footage, public keys, signatures, internet protocol (IP) addresses, device identifiers, metadata and so forth.

A picture may show a whole person, and very high resolution (VHR) arguably allows the identification of that person when considering, for example, that person’s height, body type and clothing. Likewise, very high resolution images could enable a person to be identified via the objects (home, cars, boats etc.) and places (location data) associated with that person [3]. The lawfulness of processing such images needs to be assured.

As **massive constellations of small satellites**²⁹ are becoming a staple in low

²⁸Regulation (EU) 2016/679 (General Data Protection Regulation) on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, OJ L 119, 04.05.2016.

²⁹The EO constellation will be centred at 600 km, which spans a large range of altitudes. It

Earth orbits (LEOs), larger influx of data, observation capabilities and high-quality imagery from EO satellites [62] is expected to become more widely available on a regular basis. EO massive constellations may provide more frequent image capture and updates (capturing a single point several times a day) at a much lower cost. Users can plan both the target and frequency, allowing more specific analysis on a particular track. Ordinarily, these collected terabytes of data must be downlinked to a ground station before being processed and reviewed. But now, enabled satellites can carry out **mission applications on board, and this includes using AI that would carry out such processing in the satellite** [35]. This means that only the most relevant data would be transmitted, not only saving on downlink costs but also allowing ground analysts to focus on the data that matters the most. For example, one company has **developed algorithms that rely on AI to analyse stacks of images and automatically detect changes, allowing users to track changes** at individual properties in any country: "This machine learning tool, it's constantly looking at the imagery and classifying things it finds within them: that's a road, that's a building, that's a flood, that's an area that's been burned" [103]. Other analytics companies feed visual data into algorithms designed to derive added value from mass images.

AI may be used, in **breach** of EU data protection and other rules, by public authorities or other private entities for mass surveillance. **VHR optical data** may have the same quality as aerial photography, and could therefore **raise privacy [7], data protection and ethical issues** [41, 96, 4, 76, 75].

In addition, **EO data** could be explored by smart video or **face recognition technologies** [14, 30] and **combined with other data streams such as the Global Positioning System (GPS), security cameras, etc.**, thus raising privacy concerns, even if the raw or pre-processed data itself does not. **We can anticipate several scenarios where the identifiability of individuals is at stake.** Applying very high resolution satellites to scan and inspect the landscape, images can be captured of buildings, cars or real estate for the purpose of showcasing, stock images, footage for publicity purposes, and such like. Those familiar with the areas captured and/or individuals in the vicinity may be able to identify those individuals and their movements and their social patterns. These are the actual risks posed by making this data available open source to be used for **any unforeseen purpose.**

The aim of the **European strategy for data** [27] is to give the EU a secure and dynamic **data-agile economy** in the world – empowering the EU with data to improve decision-making and improve the lives of all its citizens. **The EU's future**

comprises 300 non-maneuvrable 3U cubesats so is much smaller in both total areal cross-section and aggregate mass.

regulatory framework for AI aims to create an **ecosystem of trust**'. To do so, it must ensure compliance with EU rules, including rules protecting fundamental rights and consumer rights, particularly for AI systems that pose high risks, as explained in this article [26]. If a clear EU regulatory framework is required to build trust among consumers and businesses using AI in space, and therefore hasten uptake of the technology, it is necessary to **be aware of the risks of AI in space**.

While AI can do much good, it can also do **harm**. This harm might be both **material** (affecting the safety and health of individuals, including loss of life and damage to property) and **immaterial** (loss of privacy, limitations to rights including freedom of expression and human dignity, or discrimination in e.g. access to employment), and can relate to a wide variety of risks. Harnessing AI technologies to access and explore outer space and engaging in space-based commercial activities will, in all likelihood, **lead to a broad array of intended and unintended consequences** flowing from the use and misuse of such technologies, and these consequences cannot be downplayed or disregarded. However, the two most prominent and complex legal issues are considered to be privacy and data protection, and liability for erroneous positioning [97].

3.2.1 Privacy and data protection issues

The use of AI in connection with satellite imaging raises concerns relating to personal privacy and data protection. Some of the potential risks forecasted by the [25] include the following:

- ***ubiquity of “facial recognition data”*** [14]. Facial recognition data can potentially be obtained from a plethora of sources. Facial images collected and stored in a multitude of widely available databases can be used to track the movements of people through time and space. They therefore constitute a potential source for identifying individuals. Individuals may be identified via analysis of images captured by various facial recognition systems. More generally, any photograph can potentially become a piece of biometric data with more or less straightforward technical processing. Dissemination of data collected by facial recognition devices is taking place in a context of continuous self-exposure on social media, which increases people’s vulnerability to facial recognition data. A massive amount of data is technically accessible for which AI can potentially be mobilised for the purpose of facial recognition-based identification.
- ***lack of transparency***. Transparency requires that the data controller informs the data subject of the personal information collected, the purpose of

the collection, and use of the data. Transparency also entails that imagery operators inform data subjects of their right to access, correct and erase their personal data, and the procedure for exercising such rights. The transparency obligation is difficult to document, monitor and enforce given the number of different companies involved in the collection and intelligent processing of personal data.

- ***data maximisation and disproportionality of data processing.*** Space technology has a tendency towards extensive collection, aggregation and algorithmic analysis of all the available data for various reasons, which hampers the data minimisation principle. In addition, irrelevant data are also being collected and archived, undermining the storage limitation principle.
- ***lack of purpose limitation and repurposing of data.*** Since data analytics can mine stored data for new insights and find correlations between apparently disparate datasets, big data from space is susceptible to reuse for secondary unauthorised purposes, profiling, and surveillance. This undermines the purpose specification principle, which stipulates that the purpose for which the data is collected must be specified and lawful. As for repurposing, personal data should not be further processed in a way that the data subject might regard as unexpected, inappropriate or otherwise objectionable and, therefore, unconnected to the delivery of the service. Moreover, once the infrastructure is in place, facial recognition technology can easily be used for “function creep” which is the use of the technology or algorithm for purposes other than that originally intended. For instance, the purpose of VHR usage may expand to include either additional or unforeseen operations or activities compared to that originally envisaged. Function creep also describes situations when such imagery is disseminated over the Internet, which naturally increases the risk of the data being reused widely. Given these circumstances, it is difficult to ensure that the data subject can effectively control the use of the facial recognition data by giving or withholding consent.
- ***retracing.*** By analysing large amounts of data and identifying links among them, AI can be used to retrace and deanonymise data about persons [26], thereby creating new personal data protection risks even with datasets that do not include personal data per se.
- ***lack of rights of access, correction and erasure.*** Results obtained from data analysis may not be representative or accurate if the sources of the data are not subject to proper validation. For instance, AI analysis combining

online social media resources are not necessarily representative of the whole population. Moreover, machine learning may contain hidden biases in its programming or software, which can lead to inaccurate predictions and inappropriate profiling of persons. Hence, AI interpretation of data collected by high-resolution images need human validation to ensure the trustworthiness of a given interpretation and avoid an incorrect image interpretation. At best, satellite images are interpretations of conditions on Earth – a “snapshot” derived from algorithms that calculate how the raw data are defined and visualized’ [45]. This can create a black box, making it difficult to know when or why the algorithm gets it wrong. For example, one recently developed algorithm was designed to identify artillery craters on satellite images – but the algorithm also identified locations that looked similar to craters but were not craters. This demonstrates the need for metrics to assist in formulating an accurate interpretation of big space data.

- ***potential identification of individuals.*** If footage taken via VHR imaging only shows the top of a person’s head and one cannot identify that person without using sophisticated means, it is not personal data. However, if the same image was taken with the backyard of a house in view using additional imaging analytical algorithms that may enable the house and/or the owner to be identified, then that footage would be considered to be personal data. Thus, personal data is very much context-dependent. The situation escalates with the advance of “ultra-high” definition images being published online by commercial satellite companies, and the subsequent application of big data analytics tools. It might be possible to identify an individual indirectly (and show the individual’s house etc.), when high-resolution images are combined with other spatial and non-spatial datasets. Thus, while footage of people may be restricted to “the tops of people’s heads”, once these images are contextualised by particular landmarks or other information, individuals may become identifiable. *Combination of publicly available data pools with high resolution image data coupled with the integration and analysis capabilities of modern GIS [Geographic Information Systems] providing geographic keys[,] such as longitude and latitude, can result in a technological invasion of personal privacy”* [12].
- ***erosion of anonymity in the public space*** [14]. Erosion of anonymity by public authorities or private organisations is likely to jeopardise some of the fundamental privacy principles established by the GDPR. Facial recognition in public areas can end up making harmless behaviour look suspicious. Wearing a hood, sunglasses or a cap, or looking at one’s telephone or the ground

can have an impact on the effectiveness of facial recognition devices, and such behaviour can be the basis for suspicion [14]. Additionally, the interface between facial recognition systems and satellite imaging creates an opportunity for an unprecedented level of surveillance, whether by a governmental or private entity. It is not inconceivable that coupling satellite imagery with facial recognition software and other types of technology, such as sound capturing devices, may further increase the level of surveillance of people and places.

- ***fallible technologies producing unfair biases [8] and outcomes*** [14, 49]. Like any biometric processing, facial recognition is based on statistical estimates of a match between the elements being compared. It is therefore inherently fallible because it is a match based on probability. Furthermore, as the French data protection law explains, the biometric templates are always different depending on the conditions under which they are calculated (lighting, angle, image quality, resolution of the face image etc.). Every device therefore exhibits variable performance according, on the one hand, to its aims, and, on the other hand, to the conditions involved in collecting images of the faces to be compared. Space AI devices embedded with facial recognition technology can thus lead to "false positives" (a person is wrongly identified) and "false negatives" (the system does not recognise a person who ought to be recognised). Depending on the quality and configuration of the device, the rate of false positives and false negatives may vary. The model's result may be incorrect or discriminatory if the training data renders a biased picture of reality, or if it has no relevance to the area in question. Such use of personal data would be in contravention of the fairness principle.
- ***lack of transparency and (in)visibility***. This risk applies when individuals on the ground may not know that VHR satellites are in operation, or if they do, may be unsure about who is operating them and the purpose for which they are being used, which somehow causes discomfort.
- ***seamless and ubiquitous processing***. VHR combined with facial recognition technologies allows remote, contactless data processing [14]. Such a "contactless" system means that processing devices are excluded from the user's field of vision. It allows remote processing of data without people's knowledge, and without any interaction or relationship with those persons. In this scenario, data controllers need to declare the data subject's rights and the procedures for exercising these rights (Articles 13(2)(b) of the GDPR).
- ***loss of privacy and non-public areas***. Using AI with satellite imaging

presents issues about loss of control over one's personal information and activities [55, p. 70-72.][83, p. 24-29], which encompasses the right of individuals to move in their own home (including yards and gardens) and/or other non-public places without being identified, tracked or monitored [64, p. 16.].

- ***loss of privacy of association.*** This refers to people's freedom to associate with others [64, p. 16]. It is related also to the fact that footage might indicate, for example, the number of adults living in a house (based on the number of vehicles) or provide clues as to their relationships. Satellite imaging and AI provide an opportunity to ascertain and/or monitor personal associations.
- ***lack of means to verify compliance.*** The specific characteristics of many AI technologies, including opacity (the black-box effect'), complexity, unpredictability and partially autonomous behaviour, may make it hard to verify compliance with rules of existing EU law intended to protect fundamental rights, and may hamper the effective enforcement of those rules. Enforcement authorities and affected persons may lack the means to verify how a decision was made if it involved AI in space and, therefore, whether the relevant rules were respected. Individuals and legal entities may face difficulties obtaining effective access to justice in situations where such decisions can affect them negatively.

3.2.2 Ethical issues

The use of AI in connection with satellite imaging raises the following ethical issues:

- ***discrimination.*** Profiling consists of "pattern recognition, comparable to categorization, generalization and stereotyping" [37]. VHR satellite imaging combined with analytic technologies can lead to discriminatory profiling [21]. Also, satellite-based VHR may be used more on certain populations or areas where people are less likely to be able to effectively voice or act upon such concerns (i.e. marginalised populations or areas). With the use of ML and data mining, individuals may be clustered according to generic behaviours, preferences and other characteristics without necessarily being identified [20]. Profiling ultimately involves creating derived or inferred data and occasionally leads to incorrect and biased decisions (based on discriminatory, erroneous and unjustified judgements about, for instance, their behaviour, health, creditworthiness, recruitment potential, insurance risks etc.) [Edwards, 2016].
- ***public dissatisfaction.*** People could become disillusioned with surveillance and use of imagery based on the possibility that these activities can compromise

privacy and data protection rights or due to a feeling that they are being “overrun” by such technologies.

- *chilling effect*. There are situations where individuals may be unsure if they are being observed (even if there are no VHR satellites processing data about them), and they attempt to adjust their behaviour accordingly [64, p. 16].
- *imbalance*. In one prospective scenario, space technologies might produce situations of imbalance where data subjects are not aware of the fundamental elements of data processing and related consequences and are unable to negotiate what information may be kept about them and for what purpose, which has the side effect of enhanced information asymmetry. Even exercising the right to be forgotten seems difficult. Images captured for use in Google Street View may contain sensitive information about people who are not aware that they are being observed and photographed [Holdren and Lander, 2014].

If these risks materialise, the lack of clear requirements and the characteristics of space-based AI technologies make it difficult to trace potentially problematic decisions made with the involvement of AI systems. This in turn may make it difficult for persons who have suffered harm to obtain compensation under current EU and national liability legislation [25].

4 Limitations of space treaties in determining the law applicable to intelligent systems and services

Limitations naturally exist in the space law treaty regime because it employs broad legal principles accompanied by ambiguous terms and provisions. Moreover, the regime does not sufficiently reflect how access to and use of outer space has metamorphosed due to the escalation and diversification of space activities engaged in by private actors and other non-governmental entities, and due to technological advancements such as AI. The lack of international standardisation in the space law treaty regime comes to light generally when some sort of unforeseen event occurs, such as i) damage to a space asset³⁰ or ii) an act that increases the hazards of access to and use of outer space.³¹ This section will analyse and discuss limitations

³⁰Injury or harm is not limited to physical collision with a space object but also includes conduct such as jamming a satellite transmission, hijacking a satellite signal, or seizing command and control of a space object.

³¹The creation of space debris by testing an anti-satellite weapon is an example of an act that increases the hazards of access and use of space.

associated with the State-centric space legal regime, and will discuss jurisdiction and choice of substantive law issues related to the restrictive State-centric space law treaty regime.

4.1 State-centric space legal regime

The **space law treaty regime does not impose any direct obligation on non-governmental entities**. Instead, it puts all the responsibilities and obligations on only one class of space actor, the State. For instance, Article VI of the Outer Space Treaty establishes that the outer space activities of non-governmental entities are subject to restraint and control by States and not the treaty regime directly (provided that the non-governmental actor's space activity does not involve piracy, genocide or any other recognised international crimes).

The most fundamental limitation is apparent in Article XIII of the Outer Space Treaty, which recognises that the treaty provisions apply only to the activities of States, albeit including international governmental organisations and related entities. Limiting the obligations and remedies associated with space activities to States essentially **relegates the space law treaty regime to the rule of politics rather than the rule of law**. As the space economy matures, it will become necessary to have a space law legal regime directly applicable to all space actors and rooted in the rule of law rather than politics. Until international space law goes beyond the State-centric space legal regime, it will suffer from the limitation of restricted direct application and other limitations in areas such as jurisdiction and choice of applicable substantive law.

4.2 Jurisdictional limitation

There is no international body that has the jurisdiction to adjudicate space-based disputes between States and bind the States to its judgments. International jurisdiction over space-based disputes depends on the consent of all the States that are parties to the dispute [29]. Moreover, since the space law treaties do not impose direct obligations and duties on non-governmental entities, there is no basis for international jurisdiction over a non-governmental space actor, provided that the non-governmental actor's space activity does not involve piracy, genocide or any other recognised international crime.

The jurisdictional limitations of the space law treaty regime is also apparent whenever a private individual or non-governmental entity wants to pursue a remedy directly for harm caused by some space activity. In such circumstances, jurisdiction is determined by State law unless the parties to the dispute consent to private arbi-

tration. Although space law precludes a State from exercising sovereignty in outer space, space law incorporates international law that recognises a State's power to exercise jurisdiction over extraterritorial acts under certain circumstances.³² Accordingly, State law governing jurisdiction can arguably extend to AI-related disputes arising in space. Moreover, a State can enact specific legislation granting its courts or agencies jurisdiction over AI-based disputes arising from space activities. Either way, jurisdiction is properly determined only if such legislation satisfies one of the five grounds for extraterritorial jurisdiction.³³

An example of a State extending its jurisdiction to the space context is when the United States enacted a statute criminalising interference with the operation of a satellite i.e. any activity that "intentionally or maliciously interferes with the authorized operation of a communications or weather satellite or obstructs or hinders any satellite transmission."³⁴ Noticeably, the statute did not: 1) limit its application to a satellite launched by or registered in the United States, or 2) limit its application to situations affecting the national security of the United States, citizens of the United States, the economic interests of the United States or any other interest pertaining to or connected with the United States [48]. In the absence of universal jurisdiction over interference with satellites, the statute's jurisdictional scope appears to be overbroad and extends beyond jurisdictional reach consistent with international law.

In any event, there is no international harmonisation or standardisation on matters of jurisdiction over AI-related disputes arising from space activities that do not cause damage as defined by the Liability Convention, or covering situations when a remedy is being directly pursued by a private person or non-governmental entity.

4.3 Limitations of space law

The space treaties do not address the use of AI, and there is no international treaty regulating AI in space. This means that domestic legislation must be the principal source of substantive law on the use of AI in space. The lack of international regulation of AI **potentially poses complex problems relating to the applicable substantive law in disputes involving the use of AI in space.** For

³²United States v. Ali, 885 F.Supp.2d 17, 25-26 rev. in part on other grounds 885 F.Supp.2d 55 (D.D.C. 2012). Customary international law generally recognises five tenets for a State exercising jurisdiction over the extraterritorial conduct of a non-governmental entity. As the United States judiciary has recognised in the context of the United States v. Ali piracy case, the five tenets are: territorial jurisdiction, protective jurisdiction, national jurisdiction, passive personality jurisdiction, and universal jurisdiction.

³³See Id.

³⁴18 U.S.C. § 1367.

instance, if the use of AI or an intelligent space object causes damage to another space object which is cognisable under the Liability Convention, it is unclear what substantive law applies when it comes to important issues for resolving the merits of the claim, issues such as the appropriate standard of care and what constitutes fault. Article XVIII of the Liability Convention stipulates that the “Claims Commission shall decide the merits of the claim for compensation and determine the amount of compensation payable, if any.” However, neither Article XVIII nor any other provision of the Liability Convention indicates which substantive law should be used to decide the merits of the claim and determine the compensation issue. Is the appropriate substantive law the domestic law of: 1) the launching State of the space object causing the damage, 2) the State where the space object causing the damage was registered, 3) the State that owned the damaged space object or whose national owned the object, 4) the State where the damaged space object was registered, 5) the home State of the software developer who created the AI used by the space object that caused the damage? Or is it the substantive law formulated by the Claims Commission?³⁵

There is a similar choice of substantive law problem when the dispute involves a space-based injury that is not subject to the Liability Convention or when an injured non-governmental entity decides to pursue a claim directly for injury caused by space activity. If such a claim is brought to the State’s judiciary, then that State’s conflict of law provisions may help determine which substantive law applies. It is an open question whether the Liability Convention’s liability scheme for allocating fault can apply to private persons or non-governmental entities. The judiciary in Belgium and the United States have both adopted customary international law principles embodied in an international treaty as the substantive law for resolving disputes between two private parties arising in the international arena of the high seas.³⁶ Thus, the Liability Convention’s fault allocation scheme may conceivably be used in proceedings for space-based damage or harm arising from the use of AI. However, that does not eliminate the choice of law dilemma when it comes to determining causation, fault, and other merit-related issues.

As we have seen, the lack of substantive law at the international level limits the ability of the space law treaty regime to establish a harmonious or uniform legal standard for deciding claims involving AI-related space-based damage or harm.

³⁵Liability Convention Article XVI(3) allows the Claims Commission to determine its own procedure, which should include how it chooses the applicable substantive law.

³⁶*Castle John v. NV Mabeco*, 77 ILR 537 (Belgium Court of Cassation 1986); *Institute of Cetacean Research v. Sea Shepherd Conservation Society*, 708 F.3d 1099 (9th Cir. 2013). The two cases involved plaintiffs seeking injunctive and declaratory relief against a non-State actor for conduct alleged to constitute piracy under international law.

Given that the Liability Convention employs fault-based liability for extraterrestrial damage caused by a space object, it is sensible and practical that the same liability scheme should be applicable to a legal action involving extraterrestrial harm attributable to AI outside the scope of the Liability Convention, or where a non-governmental entity is a party. Otherwise, there will not be any international standard for, among other things, allocating fault and determining liability for extraterrestrial harm arising from space activities. The lack of international standardisation means that the plethora of potential substantive law choices becomes a critical issue. State laws can vary based on whether the State is a common law jurisdiction like the United States and Great Britain, a civil law jurisdiction as in most European States, a jurisdiction based on Islamic law, a State that practises some form of Marxism, or a State that has some other political or legal system.

Selecting from the buffet of substantive law choices in matters involving AI is complicated by the fundamental conceit that the space law treaty regime, like all State legal systems, is based on controlling and regulating the decisions, acts, errors, and omissions of a person or people even if made in the guise of a juridical person. AI is machine conduct. The fundamental distinction between AI conduct and human conduct is an issue that is currently facing the legal systems of technologically advanced States.

The United States is a technologically advanced State that is struggling to find the right “fit” for legal actions arising from an event involving AI. Generally, in the United States, legal actions seeking compensation for harm caused by a device or machine either claim negligence on the part of the owner/operator or are based on a theory of product liability [32]. However, both theories require that fault should be determined based on human conduct. Negligence requires human involvement. Product liability concerns a defect in software design or manufacturing, and failure to provide a warning of reasonable foreseeable injury [13, 89]. According to these studies, a design defect occurs when a foreseeable risk of harm exists and the designer could have avoided or reduced the risk by using a reasonable alternative design; a manufacturing design fault occurs when a product is not produced according to specifications; and failure to warn occurs when the responsible party fails to “provide instructions regarding how to safely use the software” [89].

Liability for an AI design defect can incur either strict liability or fault liability depending on which particular industry is involved or what kind of application is using the AI [Sword, 2019]. However, the EC’s White Paper on AI [26], seems to adopt the fault-based approach to injury caused by an AI system. In suggesting that product liability law may not be a “good fit” for AI-related injury, the White Paper recognises that “it may be difficult to prove that there is a defect in the product, the damage that has occurred and the causal link between the two”. It further notes

that “there is some uncertainty about how and to what extent the Product Liability Directive applies in the case of certain types of defects, for example if these result from weaknesses in the cybersecurity of the product”.

Nevertheless, since strict and fault liability are predicated on human conduct, a new perspective is emerging that perhaps a new liability scheme is needed for AI. Two new liability concepts proposed for AI are some form of legal personhood specifically for AI and “robot common sense” as a substitute for the “reasonable man” standard used in current jurisprudence [32]. There is also a suggestion that agency law should be used in connection with AI systems since the autonomous machine is actually an agent of the owner or operator. Regardless of how necessary it is to have a new liability standard for AI, especially in the context of outer space, any such development will, in all likelihood, have to emerge from the domestic law of States. This is a substantial limitation in the space law treaty regime for the development of uniform substantive law governing fault liability for space-based damage caused by a space object using AI.

5 Elements of legal methodology for determining the law applicable to intelligent systems and services

Legal methodology is a way of reaching a legal result in a coherent and deductive way [77]. The most common legal methodology involves a three-pronged approach consisting of 1) a method of description, 2) a method of conceptual analysis and 3) a method of evaluation [42]. This section will examine AI in space in the light of these three prongs.

5.1 Method of description

The “method of description” describes the state of affairs as it exists at present [42, p. 2.]. The previous section articulated the current status on matters of jurisdiction and choice of law relating to disputes involving the use of AI in space.

5.2 Method of conceptual analysis

The “method of conceptual analysis” concerns an abstract idea or theory and usually involves: “(1) analysis of the existing conceptual framework of and about law; (2) construction of new conceptual frameworks with accompanying terminologies”.³⁷ Since the previous section focused on the existing conceptual framework and its

³⁷[42] quoting [88].

limitations, this section will examine a new conceptual framework for determining the applicable jurisdictional and substantive law in space-based disputes caused by an AI system, at least when one party to the dispute is an EU Member State or entity. The approach would be consistent with the EC's White Paper on AI [26], which recognises the need to avoid fragmented and divergent national rules by juridical entities of its Member States on the use of AI within its markets.

The ruling delivered by the Court of Justice of the European Union (CJEU) on December 19, 2019, regarding the online accommodation-sharing platform Airbnb³⁸ could serve as an important benchmark for establishing jurisdictional boundaries. It may provide important tangential clues as to how to extend the existing frames of reference used to determine the applicable legal regimes governing **AI systems and services in space**. **Although the CJEU ruling pertains exclusively to terrestrial online platforms providing consumer-related intermediation services**, the extent of which falls considerably below the scope of in-orbit servicing, the extrapolation of this judgment to extraterrestrial matters potentially provides an alternative route to determining the legal regime for service providers and the larger issue of territoriality and State jurisdiction in space.³⁹

5.2.1 Overview of the solution adopted by the Court of Justice of the European Union

The CJEU provides a legal characterisation of Airbnb's activities which is in accordance with the recommendations set out in the Directive on Electronic Commerce (the e-Commerce Directive), concluding that the Airbnb *platform* fits the definition of *information society services*" provided in Article 2(a) of the e-Commerce Directive.

In attempting to characterise Airbnb's service offering, the Court carried out a comprehensive examination of Airbnb's online marketplace as well as its wider business model. It ultimately found that the defendant's digital platform provides a

³⁸Aff. C-390/18, YA and Airbnb Ireland UC versus Hotelière Turenne SAS et Association pour un hébergement et un tourisme professionnel (AHTOP) et Valhotel, Concl. Maciej Szpunar, Press Release n°162/19.

³⁹In the legal context of the European Union, this judgment is particularly significant in that it achieves this result without attempting to redefine the distribution of powers (as was done by the Lisbon Treaty) where space activities are concerned. Indeed, space as well as the attendant technological research & development activities remain fully entrenched within the realm and jurisdiction of shared competence between the Union and its Members States, in complete agreement with Article 4 of the Treaty on the Functioning of the European Union, which specifies that space is an area over which "the Union shall have competence to carry out activities, in particular to define and implement programmes; however, the exercise of that competence shall not result in Member States being prevented from exercising theirs".

direct intermediation service supplied remotely via electronic means, linking potential tenants and landlords, and offering to facilitate their entering into contractual agreements about future transactions.

Although the Airbnb ruling seems *prima facie* far removed from the realms of space and **AI systems and services**, it offers, nonetheless, a potentially new framework from which to examine broader issues of jurisdiction and extraterritoriality, and determine appropriate legal regimes for novel phenomena spawned by emerging technologies that elude regulatory oversight.

In this context, the CJEU's ruling of December 19, 2019, on Airbnb is a landmark decision that could prompt new insights into how **jurisdictional conflicts may be litigated in the future**. In particular, the decision raises two distinct avenues for inquiry, namely: i) its characterisation of the platform as a neutral vehicle (a delivery system) devoid of inherent liability and unconnected to jurisdiction; and ii) the territorialisation of the service delivered. In other words, the legal regime that is applicable to a platform — whether that platform is deployed on Earth or in outer space — is pegged on the content (the purveyor and/or beneficiary of the services) rather than the container (the physical platform and its operator).

(a) Emphasis on the nature of the service provided

What appears to be important to the Court is not so much the features or functionality of the platform used but the nature of the service provided. What it regards as constituting an information society service is really the purpose of that service: putting potential tenants in contact, in return for payment, with professional or non-professional landlords offering short-term accommodation services, so that tenants can book accommodation.

The fact that this service is provided by means of an electronic platform seems to be less important than the fact that it is provided at a distance, albeit by electronic means, or that it is provided at an individual's request on the basis of an advertisement disseminated by the landlord and an individual request from the tenant interested in the advertisement.

The fact that this service is provided by means of an electronic platform seems to be less important than the fact that it is provided at a distance, albeit by electronic means, or that it is provided at an individual's request on the basis of an advertisement disseminated by the landlord and an individual request from the tenant interested in the advertisement.

The platform only appears in the Court's reasoning as technical support for the service, and the main characteristic of that technical support, as far as the Court is concerned, is that it is provided remotely. The Court of Justice of the European Union notes that the service is provided "*by means of an electronic platform*" (ğ47), although it acknowledges that the technical support plays an essential role in the

provision of the service, noting that the parties come into contact only through the electronic platform of the same name (ğ47).

(b) Unbundling of intermediation and hosting services

This approach is all the more interesting in that the reasoning of the Court of Justice of the European Union concerning the Airbnb electronic platform articulates a second argument: the intermediation service provided by Airbnb by means of the eponymous electronic platform must be disassociated from the real estate transaction “*in so far as it does not consist solely in the immediate provision of accommodation*” (ğ54). In the Court’s view, it consists more in making available on an electronic platform “*a structured list of short-term accommodation (...) corresponding to the criteria adopted by persons seeking short-term accommodation*”, so that that the service (and hence the platform itself) is regarded only as “*an instrument facilitating the conclusion of contracts relating to future transactions*” (ğ53).

As the Court points out, “*it is the creation of such a list for the benefit of both guests with accommodation for rent and those seeking such accommodation which is the key feature of the electronic platform managed by Airbnb Ireland*” (ğ53).

Put differently by the same Court, the service provided by Airbnb Ireland by means of its electronic platform “*cannot be regarded as merely ancillary to an overall service falling within a different legal classification, namely the provision of accommodation*” (ğ54). Nor is it indispensable to the provision of accommodation, since this is provided directly by the landlords, whether professional or non-professional. It only provides one more channel, in addition to other ways and means, for the parties to the accommodation contract to meet and conclude the contract.

By recognising its independence, the Court renders Airbnb a service providing additional support, which serves the objectives of competition and, consequently, the market, especially since the electronic platform does not intervene in determining the price of accommodation. It is merely a means of facilitation, which includes all associated services (photographs of the asset rented, an optional instrument for estimating the rental price in relation to market averages calculated by the platform, a rating system for landlords and tenants), considered as part of “*the collaborative logic inherent in intermediation platforms, which allows, on the one hand, housing applicants to make a fully informed choice from among the housing offers proposed by landlords on the platform and, on the other hand, allows landlords to be fully informed about the seriousness of the tenants with whom they are likely to engage*” (ğ60).

(c) Consecration of the law of the country of establishment of the service provider

By classifying the intermediation service provided by the Airbnb platform as an information society service, the Court of Justice of the European Union makes

it subject to the aforementioned Directive 2000/31. This means that, “*in order to ensure effectively the freedom to provide services and legal certainty for service providers and their recipients, such information society services must in principle be subject to the legal regime of the Member State in which the service provider is established*” (Recital 22).

The attachment of an activity, whether terrestrial or space-based, to the jurisdiction of a State implies submission of that activity to the legal system of that State. According to the logic of the internal market of the European Union, the activity may be linked to a particular State that is “*the State in which the service provider is established*”. Indeed, since the legal orders of each Member State are supposed to integrate the provisions of the regulations or directives of the European Parliament and the Council, these legal orders are made up of harmonised legislative or regulatory texts.

This is all the more true since the principle of the primacy of Community law gives precedence to European rules over national law and since this particular European rule is itself directly applicable. An European citizen can therefore ensure that it will be applied by the national court whether or not the national law is in conformity with European law.

This is why, in the logic of European integration, the principle adopted to determine which law is applicable to a service activity is that it must be the law of the country in which the service provider is established or, in the case of broadcasting by means of satellite systems, the law of the country in which the signal is transmitted.

Comparable reasoning can be articulated with regard to a platform deployed in space. Application of the law of the country of the origin of the service provided by means of an intelligent space platform is preferable, in our view, to applying the law of the country of consumption of the service.

(d) Obligation of prior notification of national provisions

The Airbnb decision is also interesting in that it obliges Member States to notify the European Commission when their national legislation is more restrictive than the EU legislation. This is an interesting idea that can be transposed to international relations. Moreover, a comparable practice exists in the air transport sector, which leaves States sovereign over their respective airspaces and, in the name of this sovereignty, allows them to have differences between their national legislation and international rules known to the International Civil Aviation Organization (ICAO). These differences are accepted and respected on the condition that they have been notified to the ICAO. The same mechanism could be transposed to space law.

In the Airbnb judgment, the European Court of Justice did not proceed differently. It set aside the national law of the country of consumption of service, in this case French law, on two separate grounds. The first arises from the principle of the

free movement of information society services between Member States, which the Court of Justice considers to be one of the objectives of Directive 2000/31, going so far as to point out that “this objective is pursued by means of a mechanism for monitoring measures liable to undermine it” (ğ91). The second is a corollary of the first, since it follows from the obligation imposed on Member States, by Directive 2000/31, to notify the Commission of measures restricting or liable to restrict the free movement of information society services prior to their entry into force.

The Court pointed out that the obligation to notify is not “*a mere information requirement*”, but corresponds in fact to “*a procedural requirement of a substantive nature justifying the non-applicability to individuals of non-notified measures restricting the free movement of information society services*” (ğ94). As the Court also pointed out, this is indeed “*a standstill obligation on the part of the State intending to adopt a measure restricting the freedom to provide an information society service*” (ğ93).

In its judgment of 19 December 2019, the Court of Justice of the European Union did not reject this eventuality. On the contrary, it recognised that, in extending the provisions of Article 3(4) of Directive 2000/31, Member States have the option of taking measures that derogate from the principle of the free movement of information society services with regard to a given information society service falling within a relevant field. However, in addition to the procedural obligation to notify referred to above, it laid down three substantive conditions which must be satisfied (ğ84):

- the restrictive measure concerned must be necessary in order to guarantee public policy, protect public health, ensure public security or protect the consumer AND,
- it must be taken against an information society service that effectively undermines or constitutes a serious and grave risk of undermining these objectives, AND,
- it must be proportionate to these objectives.

(e) Key takeaways from the Airbnb case

From a space law perspective, this ruling is especially significant as its rationale may be extended to space platforms that are assembled in outer space and that are used to provide AI systems and services in space and whose connections to terrestrial jurisdictions are inconclusive.

Such stations and platforms, regardless of their complexity, purpose and functionality, remain in essence supporting tools and instruments designed to facilitate the provision of a service. In other words, they are a means to an end. As such, the ruling of the CJEU would retain its role as a deciding factor to help courts determine the true nature and scope of an information communication technology related service, whether on Earth or in outer space. In other words, when examining the concept of platform, the Court excluded all metonymic reasoning and retained that it is the nature of the service that is being provided via that platform that is

the primary consideration in making a legal characterisation, not the characteristics of the platform as a vehicle for delivering that service. Further, the ruling argues that even in circumstances where the provision of a service must in fine be conflated with the medium that is used to deliver it⁴⁰, the provision of the service must be considered over the medium that is used to deliver it, which must be considered a secondary aspect.

5.3 Method of evaluation

The “method of evaluation” involves examining “whether rules work in practice, or whether they are in accordance with desirable moral, political, economic aims, or, in comparative law, whether a certain harmonisation proposal could work, taking into account other important divergences in the legal systems concerned”.⁴¹ As discussed below, the Airbnb case may be applicable to disputes involving AI systems and services in space.

A similar rationale can be equally replicated in the context of AI systems and services in space **supplied** via outer space platforms powered by AI technologies. Drawing on a concept that has been common to both telecommunications law and electronic communications law ever since those industries opened up to market competition, the support service delivered by a given platform is in and of itself a bearer service (or data service) that makes an infrastructure available to users. Such a service must be distinguished from the global service provision supplied through the platform-as-a-medium. As the service becomes increasingly dematerialised, it too becomes progressively disassociated from its medium.

In the case of in-orbit servicing, if the bulk of that service is actually delivered in orbit⁴², the foreseeable legal challenge lies in accurately identifying the substance and nature of this service provision and defining its governing legal regime. This must apply even where the service provision merges with its delivery platform to such an extent that a platform governed by artificial intelligence becomes materially indivisible from the services that it is designed to deliver. From a legal perspective, the delivery of such a service calls for a distinct characterisation that falls under the authority of the principles governing the activities of States in the exploration and use of outer space, including the moon and other celestial bodies (i.e. the principles of the Outer Space Treaty).

⁴⁰Likely because the contracting parties could only establish contact through the intermediation of this service/tool.

⁴¹[42] quoting [38, p. v.]

⁴²Be it remote computation, temperature-controlled storage, maintenance operations, rescue missions, remote sensing and Earth observation or big data storage, etc.

This particular requirement raises the larger question of how to define the boundaries of the legal forum and the jurisdictional competence of the State i.e. the range of the applicability of national laws over matters beyond the traditional purview of national legislation. The CJEU's judgment of December 19, 2019, on the matter of Airbnb's digital platform offers an important contribution to this question as well. In characterising the intermediation service delivered by the Airbnb digital platform as an information society service, the CJEU places the defendant's business under the scope of Directive 2000/31. The directive lays down that "*in order to improve mutual trust between Member States [and] to effectively guarantee freedom to provide services and legal certainty for suppliers and recipients of services, such information society services should in principle be subject to the law of the Member State in which the service provider is established*".⁴³

Beyond the basic principles of the freedom to provide services and legal certainty⁴⁴, binding in-orbit service delivery provided via a smart space platform to the legal jurisdictional authority and to the legal regime of a particular State provides a fresh outlook on the leading doctrine established by Article VIII of the OST.⁴⁵ Where weighing the various connecting factors enables an Earth or space-bound activity to be ascribed a given national jurisdiction, the activity is bound by the legal regime of that State. Therefore, in keeping with the internal market rationale of the EU, jurisdiction can be established on the basis of *the State in which the service provider is domiciled* (see [18]).

Within the particular context of EU law, considering that the legal regime of each Member State is required to integrate the statutory and regulatory provisions issued by the European Parliament and the Council, all the national legislative and regulatory frameworks are supposed to be harmonised across EU jurisdictions, in compliance with the overarching principle of the primacy of Community law.⁴⁶ This principle requires that the EU rule of law must always prevail over national law where there is a conflict of laws, and that EU regulations have direct application within national jurisdictions. As a result, all EU citizens have the prerogative to avail themselves of the right to petition a national court to enforce the application of an EU statutory or regulatory provision over national law, regardless of whether the national law is compliant with EU legislation.

⁴³Directive 2000/31/CE, Recital nr22.

⁴⁴Which are not unique to the internal European Union market, and which apply in equal measure to terrestrial commerce as to outer space commerce.

⁴⁵With its twofold implication of jurisdictional boundary and government control.

⁴⁶Declaration 17 concerning primacy in Declarations annexed to the Final Act of the Intergovernmental Conference which adopted the Treaty of Lisbon (December, 13, 2007). See also the consolidated protocols, annexes and declarations attached to the treaties of the European Union.

As is consistent with the guiding principles of European integration, the basic legal principle used when determining the appropriate legal regime applicable to the provision of a commercial service tethers the legal forum to the service provider's place of establishment, or, in the case of satellite frequency broadcasting, to the law of the country from which the signal is emitted. For our purposes, the latter criterion can be a particularly helpful connecting factor where, in the case of sophisticated information and communications technology (ICT) semi-autonomous applications developed by international teams cutting across traditional jurisdictions, the original place of establishment cannot be conclusively determined.

In the light of the preceding discussion, one proposed way out of the jurisdictional quandary raised by emerging intelligent technologies in outer space would be to bind the provision of the service to the legislation and to the jurisdictional forum of the beneficiary of that service i.e. the customer or the consumer of that service. Such an approach would have the advantage of bringing an added level of clarity to determine the appropriate legal forum and address the lingering difficulty of establishing clear connecting factors that bind orbital operators to terrestrial jurisdictions.

The latter situation arises when the country of registration is designated as the sole applicable jurisdiction where the "customer" is an actual space object that is subject to mandatory registration. This particular situation brings many challenges in the context of the intersection of the digital economy and the space industry. Rethinking liability around service users and service purveyors might be a way forward that is more in line with the direction that the industry seems to be taking as a whole.

The proposed solution might also put a stop to the growing practice of many States, due to the difficulties of tracking and controlling the activities of private operators in space, of starting to "[*relax*] the registration and supervision of corporations [*incurring the risk*] of possible liability" and failing to respect the duty of care imposed by the treaties [104].

Going forward, the applicable legislation could be the national jurisdiction of the natural or legal entity that benefits either directly or indirectly from the service that is being supplied in orbit.

Such a solution would usher in an unprecedented level of transparency and legal certainty to all the stakeholders involved, and would further benefit from existing legal scholarship and regulatory frameworks that are already in place in other areas of international law as well as regulations to streamline oversight mechanisms while also stimulating industrial development. For instance, with particular regard to State subsidies and EU community State aid rules, there is ample legal expertise and established jurisprudence to help lawmakers trace international finance networks to determine State accountability and expose hollow intermediaries [66].

Finally, such a solution would provide the flexibility required to enable any concerned State to introduce appropriate oversight and control mechanisms with its own legislation. This aligns with the observable trend of States relying ever more on national legislation rather than international consensus to regulate competitive markets.

6 AI techniques to support space law

The triad of “law, space, and AI” is an instance of the “law, science, and technology” triad. The intersection between the elements of the latter triad is quite extensive and goes beyond the scope of this whole handbook. The relationship between the first triad and this article is fairly similar. There are several open questions and developments in AI that concern the law and legal compliance. Those that concern the legal questions raised in this article so far are, of course, relevant. For instance, the developments of machine learning (federated learning, transfer learning, generative adversarial networks) or real-time analysis with big data in general are approaches that might facilitate compliance with the General Data Protection Regulation and are relevant for space applications too. The big question of liability related to automated decision-making or machine learning algorithms, or the possible accountability of autonomous agents, are also open questions for law. For now, answering these questions means relying on State law and general principles of international law where space law is inapplicable or uncertain. Space is an area where legal and technical solutions to the unresolved issues are of great significance because of our increasing reliance on AI for space activities. Autonomous space agents will need to reason about legal obligations under applicable law when making decisions. Accordingly, the processes discussed in the Handbook of Legal AI should be applied to general questions of legal reasoning, relevant formal systems and other AI applications related to space technologies and space law. In the next section, we highlight only one approach introduced in this book that is applicable to the current and foreseeable state of space law.

6.1 Legal knowledge representation in the space AI domain

Machine-readable modelling of a consensually shared domain of space law would increase the chance that such legal knowledge will be connected across the Web and used in different applications. As discussed, the normative sources in this domain are multiple and heterogeneous, thus ontologies would appear to be the most suitable way of mapping this body of knowledge [40]. The interconnections within the space

law framework makes it a natural domain for knowledge representation, sharing and reuse.

Legal fragmentation suitable for ontologies

The space law treaty regime is complemented and supplemented by **national legislation** from each country providing thorough provisions and sufficient clarify regarding activities that are not directly addressed in the vague, imprecise, overly broadly formulated and ambiguous provisions of the Outer Space Treaty [65]. Analysis of some of the most significant space legislation immediately confirms that space law is not a **unified single text** but is a combination of **separate legal texts** (that do not all have the same legal value), **i.e. what we have is legal fragmentation**. And while we can observe an intention to keep some questions subject to international treaties and others subject entirely to national jurisdictions, these questions are highly interconnected in legal practice, as we saw above, and thoroughfares are required between supposedly separate areas.

Convergence of legal mechanisms as classes of a space law ontology

Alongside the heterogeneity of the form and contents of national legal texts — simultaneously reflecting the legal traditions of each State, their degree of involvement in the space economy and, more and more often, their willingness to differentiate themselves from other nations by offering more favourable conditions for space traders (forum shopping) — their **convergence** is reflected in the fact that the most relevant legal mechanisms are used in each different jurisdiction. The following eight provision types represent the basic schema (domain-specific classes) for any space law ontology that describes the knowledge embedded in the different legal documents:

1. authorisation and licensing
2. continuous supervision of non-governmental activity
3. liability
4. insurance
5. space debris removal
6. State strategic interest
7. registration process or registry
8. transfer of ownership in space

6.2 Relevant knowledge resources

To develop a domain-specific legal ontology, it is necessary to use the most authoritative relevant knowledge resources [57] from that specific legal domain. We have

therefore obtained both non-ontological and ontology-based resources to be further engineered.

Non-ontological resources. Non-ontological resources (NOR) in the legal domain are knowledge resources whose semantics have not been formalised yet in an ontology but which have related semantics that allows ontological interpretation of the legal knowledge they hold [74]. In fact, using non-ontological resources about the space domain that conveys consensus in the field brings certain benefits e.g. interoperability in terms of the vocabulary used, information browse/search capabilities, decrease in the knowledge acquisition bottleneck, reuse etc. The following non-exhaustive resources are recommended because they comprise highly reliable domain-related content from the websites of organisations that have knowledge of the domain. They serve as domain knowledge resources for the development of a space ontology. They are structured and semantically rich taxonomies that serve to annotate the data elements in an ontology. Reuse of these resources can enable the development of a common terminology, i.e. a harmonised high-level taxonomy of space legal concepts and terms, to characterise space law. More concretely, we suggest that the following resources can be used for automatic ontology population:

- Space Legal Tech [81], a tool representing the regulations and national space agencies of 100 countries (depicted in 1);
- the USA National Aeronautics and Space Administration (NASA) Thesaurus (accessible in machine-readable form) [54], and Taxonomy [51], documenting a high-level set of terms that can be used to map various data structures;
- the Union of Concerned Scientists (UCS) Satellite Database [91];
- glossaries, such as the European Space Agency Science’s Glossary [24] and others [52, 94, 28, 53];
- the ESA Earth Observation Knowledge Navigational System [105], a knowledge management system for EO imagery;
- A Guide to Space Law Terms [36]; and
- “Spationary” [80], a work-in-progress database with structured business and space law concepts and terms (illustrated in 2).

Ontological Resources. We leverage existing space-related ontological resources (semantically structured information about this domain) which can be reused or extended to any ontology modelling space law, which means that classes and/or instantiations from these existing ontologies can be imported.



Figure 1: National space legislation from France [81]


Concept	Common definition	Laws and Treaties	Governmental, institutional or official sources	Legal Scholarship	Related terms	Examples
International Liability	A State Party's obligation to compensate another Third State for any injury that it caused to the people or property of the latter	"Each State Party to the Treaty that launches or procures the launching of an object into outer space, including the moon and other celestial bodies, and each State Party from whose territory or facility an object is launched, is internationally liable for damage to another State Party to the Treaty or to its natural or juridical persons by such object or its component parts on the Earth, in air or in outer space[.]", Outer Space Treaty Art. VII	"It is important to note that international responsibility under Article VI [of the] Outer Space Treaty is born for 'national activities in outer space' while the matter of international liability is tied to 'space objects.' Arguably, only the latter may raise the definitional issue of whether space debris is or is not a 'space object.'", Comm. on the Peaceful Uses of Outer Space, Scientific & Tech. Subcomm. Rep. on its 48th Sess., Feb. 7-18, 2011, U.N. Doc.A/AC.105/C.1 (Feb. 3, 2011)	States are, under Article VII of the Outer Space Treaty, "liable for damage caused to another state through its own space activities or of those subject to its jurisdiction, licensing and supervision. Such an extension of responsibility and liability of a state to damage caused by its non-state entities is unusual in international law.", Francis Lyall & Paul B. Larsen, <i>Space Law: A Treatise</i> 66 (2009)	International Responsibility and Liability	Soviet Kosmos 954: Soviet Union launched a radioactive satellite in 1978 which failed and crashed in Northern Canada, and deposited debris over a wide area. The Canadian government spent over \$14M cleaning up the debris. OST + Liability Convention: Soviet Union held responsible, as the launcher state. However it only paid half of the damages  Soviet Satellite Out Of Control!

Figure 2: Excerpt of the concept of liability from Spationary [80]

In furtherance of the development of any ontological artefact, members of the space community should provide domain expertise, including verifying the accuracy of the knowledge expressed in the logical formalisations of the ontologies.

Ontological resources	Description
Ontology for space object [70]	Analysis of the category of space object and its subcategories. Space objects include artificial objects such as spacecraft, space stations, and natural space objects.
Ontology-based knowledge management for space data [71]	Discusses aspects of ontological engineering in knowledge management architectures for space data.
Ontology for Satellite Databases [72]	Offers a domain-specific terminology and knowledge model for space data systems. Where data is drawn from multiple sensors or databases, ontologies can foster information fusion via this backbone terminology.
Space Surveillance Ontology in XML Schema [63]	Captures data structures, content and semantics from a targeted military domain of space surveillance.
Orbital Debris Ontology [69]	Seeks to support orbital debris remediation by modelling orbital debris in an ontology and developing accurate and reusable debris classification.
Space Situational Awareness Ontology [73]	Domain coverage of all space objects in the orbital space environment and relevant space situational awareness (SSA) entities.
NASA Sweet Ontologies [79, 67]	A set of approximately 200 modular ontologies, collectively consisting of approximately 6,000 category terms intended to provide a knowledge base representing Earth science data and knowledge.

Ontology-based resources in the legal domain. Building on the risks for privacy and data protection described and assessed in Section 3.2, we aim to reuse ontologies that model concepts relating to the protection of personal data such as the Data Protection Ontology [5], PrivOnto [56], PrOnto [59] and GDPRtEXT (GDPR text extensions) [60]. Core legal domain ontologies, such as Eurovoc, Legal RuleML [58], and the European Legal Identifier (ELI) ontology, are also useful.

7 Conclusion

In this article, we discussed how “intelligent” systems and services raise legal problems, starting with the applicable law relating to privacy, data protection and liability. These legal challenges call for solutions that the international treaties in force cannot determine and implement sufficiently. For this reason, we suggested a legal methodology that makes it possible to link intelligent systems and services to a system of applicable rules. We also proposed legal informatics tools that could be used for the domain of space law.

References

- [1] Space Security Index. <https://spacesecurityindex.org/wp-content/uploads/2014/10/spacesecurityindexfactsheet.pdf>, 2014.
- [2] Aerospace. Artificial intelligence gets ahead of the threats, 2018. 13 December, 2018.
- [3] G. Aloisio. Privacy and data protection issues of the european union copernicus border surveillance service, 2017. Master’s thesis, University of Luxembourg.
- [4] M. S. Aranzamendi, R. Sandau, and K. Schrogl. *Current Legal Issues for Satellite Earth Observation*. European Space Policy Institute, 2010.
- [5] C. Bartolini, R. Muthuri, and C. Santos. Using ontologies to model data protection requirements in workflows. In M. Otake, S. Kurahashi, Y. Ota, K. Satoh, and D. Bekki, editors, *Proceedings of the 9th International Workshop on Juris-informatics (JURISIN, 2015)*, volume 10091 of *New Frontiers in Artificial Intelligence. JSAI-isAI 2015, Lecture Notes in Computer Science*, pages 27–40. Springer, 2015.
- [6] British Broadcasting Corporation BBC. Us lifts restrictions on more detailed satellite images. <https://www.bbc.com/news/technology-27868703>, 2014. 16 June, 2014.
- [7] C. Beam. Soon, satellites will be able to watch you everywhere all the time—Can privacy survive?, 2019. MIT Technology Review.
- [8] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the First Conference on Fairness, Accountability and Transparency, PMLR*, pages 81:77–91, 2018.
- [9] Giovanni Casini, Livio Robaldo, Leendert van der Torre and Serena Villata. *Handbook of Legal AI*, College Publications, 2022.
- [10] CBS. Google earth used to bust oregon medicinal marijuana garden, police say, 2013. CBS News. 22 October, 2013.
- [11] B. Cheng. Article vi of the 1967 space treaty revisited: ‘international responsibility’, ‘national activities’, and ‘the appropriate state’. *Journal of Space Law*, 26(1):7-32, 1998.
- [12] S. Chun and V. Atluri. Protecting privacy from continuous high-resolution satellite surveillance. In B. Thuraisingham et al., editor, *Data and Application Security*, vol-

- ume 73 of *International Federation for Information Processing*, 2002.
- [13] J. Chung and A. Zink. Hey watson - can i sue you for malpractice? examining the liability of artificial intelligence in medicine. *Asia Pacific Journal of Health Law & Ethics*, 11:51-68, 2018.
- [14] Commission nationale de l’informatique et des libertés CNIL. Facial recognition. for a debate living up to the challenges. <https://www.cnil.fr/sites/default/files/atoms/files/facial-recognition.pdf>, 2019. 15 November, 2019.
- [15] Council of the European Union. Outcome of proceedings: Proposal for a Regulation of the European Parliament and of the Council establishing the space programme of the Union and the European Union Agency for the Space Programme and repealing Regulations (eu) no 912/2010, (EU) no 1285/2013, (EU) no 377/2014 and Decision 541/2014/EU. <https://data.consilium.europa.eu/doc/document/ST-7481-2019-INIT/en/pdf>, 2019.
- [16] M. F. Cuellar. A simpler world? on pruning risks and harvesting fruits in an orchard of whispering algorithms. *University of California Davis Law Review*, November; 51:27, 2017.
- [17] A. de Concini and J. Toth. The future of the european space sector—How to leverage Europe’s technological leadership and boost investments for space ventures. https://www.eib.org/attachments/thematic/future_of_european_space_sector_en.pdf, 2019. European Investment Bank.
- [18] J. de Poulpiquet. *L’immatriculation des satellites. Recherches sur le lien de rattachement à l’Etat d’un objet lancé dans l’espace*. PhD thesis, 2018.
- [19] J. A. Dennerley. State liability for space object collisions: The proper interpretation of ‘fault’ for the purposes of international space law. *European Journal of International Law*, 29(1):281-301, 2018.
- [20] B. Van der Sloot. Privacy in the post-nsa era: Time for a fundamental revision? *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 5:2, 2014.
- [21] Machine Learning E. Denham. Big Data, Artificial Intelligence and Data Protection. Information Commissioner’s Office, UK, 2017.
- [22] The European Space Agency ESA. What is space 4.0? https://www.esa.int/About_Us/Ministerial_Council_2016/What_is_space_4.0, 2016. [accessed 4 May 2019].
- [23] The European Space Agency ESA. Automating collision avoidance. https://www.esa.int/Safety_Security/Space_Debris/Automating_collision_avoidance, 2019.
- [24] The European Space Agency ESA. Glossary. https://www.esa.int/Our_Activities/Space_Transportation/Glossary, n.d.
- [25] European Commission. Ethics guidelines for trustworthy ai. <https://data.europa.eu/doi/10.2759/346720>, 2019. Directorate-General for Communications Networks, Content and Technology, Publications Office.
- [26] European Commission. White Paper on Artificial Intelligence—A European ap-

- proach to excellence and trust. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf, 2020. COM 65 Final. European Union. 10 June, 2020.
- [27] European Commission. A European Strategy for data. Sharing Europe’s digital future. <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>, n.d.
- [28] Fact Monster. Space glossary. <https://www.factmonster.com/math-science/space/universe/space-glossary>, 2017.
- [29] M. S. Firestone. Problems in the resolution of disputes concerning damage caused in outer space. *Tulane Law Review*, 59:747-763, 1985.
- [30] European Agency for Fundamental Rights FRA. Facial recognition technology: fundamental rights considerations in the context of law enforcement. <https://fra.europa.eu/en/publication/2019/facial-recognition>, 2019.
- [31] B. Gattle. Moving from Newspace to “Nowspace”. <https://www.satellitetoday.com/innovation/2019/07/03/moving-from-newspace-to-nowspace/>, 2019.
- [32] I. Giuffrida. Liability for ai decision-making: some legal and ethical considerations. *Fordham Law Review*, 88:439, 2019.
- [33] Global Forest Watch. Amapá police use forest watcher to defend the Brazilian Amazon. <https://www.globalforestwatch.org/blog/people/amapa-police-use-forest-watcher-to-defend-the-brazilian-amazon/>, 2018.
- [34] M. A. Gray. The international crime of ecocide. *California Western International Law Journal*, 26:215, 1996.
- [35] A. Harebottle. Space 2.0: Taking ai far out. <https://interactive.satellitetoday.com/via/december-2019/space-2-0-taking-ai-far-out/>, 2019.
- [36] H. R. Hertzfeld, editor. *A Guide to Space Law Terms*. Space Policy Institute. George Washington University and Secure World Foundation, 2012.
- [37] M. Hildebrandt and S. Gutwirth. *Profiling the European Citizen*. Springer, Dordrecht, 2008.
- [38] M. Van Hoecke, editor. *Methodologies of Legal Research — Which Kind of Method for What Kind of Discipline?* Hart Publishing, Oxford and Portland, Oregon, 2011.
- [39] R. Hollingham. Google earth has given us a new way of looking at our cities and neighbourhoods—from space. richard hollingham visits the satellite factory building to see what’s coming next. <https://www.bbc.com/future/article/20140211-inside-the-google-earth-sat-lab>, 2014. BBC. 11 February, 2014.
- [40] L. Humphreys, C. Santos, L. Di Caro, G. Boella, L. van der Torre, and L. Robaldo. Mapping recitals to normative provisions in eu legislation to assist legal interpretation. In A. Rotolo, editor, *Legal Knowledge and Information System*, volume 279 of *Frontiers in Artificial Intelligence and Applications*, pages 41–49. IOS Press, 2015.
- [41] International Telecommunication Union ITU. Study group 17 at a glance. <https://www.itu.int/en/ITU-T/about/groups/Pages/sg17.aspx>, n.d.
- [42] M. Jovanovi. Legal methodology & legal research and writing - a very short introduction. [http://147.91.244.8/prof/materijali/jovmio/mei/Legal%](http://147.91.244.8/prof/materijali/jovmio/mei/Legal%20)

- 20methodology%20and%20legal%20research%20and%20writing.pdf, 2019.
- [43] C. E. A. Karnow. Liability for distributed artificial intelligences. *Berkeley Technology Law Journal*, 11:147:189–190, 1996.
- [44] W. Kowert. The foreseeability of human-artificial intelligence interactions. *Texas Law Review*, 96:181-183, 2017.
- [45] M. Laituri. Satellite imagery is revolutionizing the world. but should we always trust what we see?
<https://theconversation.com/satellite-imagery-is-revolutionizing-the-world-but-should-we-always-trust-what-we-see-95201>, 2018. The Conversation, 4 June, 2018.
- [46] G. A. Long. Small satellites and state responsibility associated with space traffic situational awareness. <https://commons.erau.edu/stm/2014/thursday/17/>, 2014. First Annual Space Traffic Management Conference “Roadmap to the Stars” at Embry-Riddle Aeronautical University, Daytona Beach, Florida, November 6, 2014.
- [47] G. A. Long. Artificial intelligence and state responsibility under the outer space treaty. In *Proceedings Of The International Institute Of Space Law*. Eleven International Publishing, 2018.
- [48] G. A. Long. Legal basis for a state’s use of police power against non-nationals to enforce its national space legislation, 2019. 70th International Astronautical Congress, Washington, D.C., 23 October, 2019.
- [49] S. Louradour and L. Madzou. White paper. a framework for responsible limits on facial recognition. use case: Flow management paper, 2020. World Economic Forum, 2020.
- [50] M. Lucas-Rhimbassen, C. Santos, G. Long, and L. Rapp. Conceptual model for a profitable return on investment from space debris as abiotic space resource, 2019. At 8TH European Conference for Aeronautics and Space Sciences (EUCASS). Held in Madrid, Spain, 1-4 July, 2019.
- [51] NASA. Nasa taxonomy 2.0. <https://www.loc.gov/item/lcwan0014329/>, 2009. Web Archive—Retrieved from the Library of Congress.
- [52] NASA. Basics of space flight glossary. nasa science solar system exploration. <https://solarsystem.nasa.gov/basics/glossary/>, n.d.
- [53] NASA. Glossary. <https://hubblesite.org/glossary>, n.d.
- [54] STI Program NASA. Nasa thesaurus. <https://sti.nasa.gov/nasa-thesaurus/>, 2012. Data file.
- [55] H. Nissenbaum. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books. Stanford, California, 2010.
- [56] A. Oltramari, D. Piraviperumal, F. Schaub, S. Wilson, S. Cherivirala, T. B. Norton, N. C. Russell, P. Story, J. Reidenberg, and N. Sadeh. Privonto: A semantic framework for the analysis of privacy policies. *Semantic Web*, Jan 1;9(2:185–203, 2018.
- [57] M. Van Opijnen and C. Santos. On the concept of relevance in legal information retrieval. In *Artificial Intelligence and Law Journal*, 25:65–87, 2017.

- [58] M. Palmirani, G. Governatori, A. Rotolo, S. Tabet, H. Boley, and A. Paschke. Legal-ruleml: Xml-based rules and norms. In F. Olken, M. Palmirani, and D. Sottara, editors, *Rule-Based Modeling and Computing on the Semantic Web, RuleML 2011*, volume 7018 of *Lecture Notes in Computer Science*, 2011.
- [59] M. Palmirani, M. Martoni, A. Rossi, C. Bartolini, and L. Robaldo. Pronto: Privacy ontology for legal reasoning. In *International Conference on Electronic Government and the Information Systems Perspective*, pages 139–152. Springer, 2018.
- [60] H. J. Pandit, K. Fatema, D. O’Sullivan, and D. Lewis. Gdprtext-gdpr as a linked data resource. In *European Semantic Web Conference 2018, Jun 3*, pages 481–495. Springer, 2018.
- [61] A. Perrazzelli and P. R. Vergano. Terminal dues under the upu convention and the gats: An overview of the rules and of their compatibility. *Fordham International Law Journal*, 23:736-747, 1999.
- [62] G. Popkin. Technology and satellite companies open up a world of data. <https://www.nature.com/articles/d41586-018-05268-w>, 2018. 29 May, 2018.
- [63] M.K. Pulvermacher, D. L. Brandsma, and J. R. Wilson. A space surveillance ontology captured in an xml schema, 2000. MITRE, Center for Air Force C2 Systems, Bedford, Massachusetts.
- [64] D. Wright R. L. Finn and M. Friedewald. *Seven types of Privacy*. Springer, Dordrecht, 2013.
- [65] L. Rapp. Space lawmaking. <https://www.thespacereview.com/article/3523/1>, 2018. The Space Review, 2 July, 2018.
- [66] L. Rapp and P. Terneyre. Lamy droit public des affaires, 2020. Lamy Kluwer no 774 et seq.
- [67] R. G. Raskin and M. J. Pan. Knowledge representation in the semantic web for earth and environmental terminology (sweet). *Computers & Geosciences*, November 1;31(9):1119–25, 2005.
- [68] R. Rosenstock and M. Kaplan. The fifty-third session of the international law commission, 2002. 96(2):412-9.
- [69] R. J. Rovetto. An ontological architecture for orbital debris data. *Earth Science Informatics*, Mar;9(1):67–82, 2016.
- [70] R. J. Rovetto. Space object ontology. <https://philarchive.org/archive/ROVS00>, 2016.
- [71] R. J. Rovetto. Ontology-based knowledge management for space data, 2017. In 68th International Astronautical Congress, Adelaide, Australia.
- [72] R. J. Rovetto. An ontology for satellite databases. *Earth Science Informatics*, Dec;10(4):417–27, 2017.
- [73] R. J. Rovetto and T. S. Kelso. Preliminaries of a space situational awareness ontology, 2016. At 26th AIAA/AAS Space Flight Mechanics meeting, Napa, California, 17 Feb 17, 2016.
- [74] C. Santos, P. Casanovas, V. Rodríguez-Doncel, and L. van der Torre. Reuse and

- reengineering of non-ontological resources in the legal domain. In *AI Approaches to the Complexity of Legal Systems AICOL*, pages 350–364. Springer, 2015.
- [75] C. Santos, D. Miramont, and L. Rapp. High resolution satellite imagery and potential identification of individuals. In P. Soille, S. Loekken, and S. Albani, editors, *Proceedings of the 2019 conference on Big Data from Space (BiDS'2019)*, pages 237–240. Publications Office of the European Union, Luxembourg, 2019.
- [76] C. Santos and L. Rapp. Satellite imagery, very high-resolution and processing-intensive image analysis: Potential risks under the GDPR. *Air and Space Law*, 1;44(3):275–295, 2019.
- [77] K. Schadbach. The benefits of comparative law: A continental european view. *Boston University International Law Journal*, 16:331, 1998.
- [78] V. Sharma. Mini satellites, maximum possibilities. <https://www.livemint.com/Leisure/yEXAK06k0UWRLtV6rzdQaP/Mini-satellites-maximum-possibilities.html>, 2018. published 10 November 2018; accessed 4 May 2019.
- [79] SIP. Official repository for Semantic Web for Earth and Environmental Terminology (SWEET) Ontologies. <https://github.com/ESIPFed/sweet>, n.d.
- [80] Space Institute for Research on Innotative Use of Satellites SIRIUS. Spationary. <https://chaire-sirius.eu/en/research/spationary>, n.d.
- [81] Space Institute for Research on Innovative Use of Satellites SIRIUS. Space Legal Tech. <https://spacelegaltech.com/>, n.d.
- [82] P. Soille, S. Loekken, and S. Albani. Proceedings of the 2019 conference on big data from space (BiDS'2019), 2019. Publications Office of the European Union, Luxembourg.
- [83] D. Solove. *Understanding Privacy*. Cambridge, Massachusetts, 2008.
- [84] L. B. Solum. Legal personhood for artificial intelligences. *North Carolina Law Review*, 70:1231-1287, 1992.
- [85] L. Soroka and K. Kurkova. Artificial intelligence and space technologies: legal, ethical and technological issues. *Advanced Space Law*, April;3(1):131–139, 2019.
- [86] Space News. Digital endeavours in space. <https://spacenews.com/digital-endeavors-in-space/>, 2019.
- [87] E. Stewart. Self-driving cars have to be safer than regular cars. the question is how much. <https://www.vox.com/recode/2019/5/17/18564501/self-driving-car-morals-safety-tesla-waymo>, 2019.
- [88] R. S. Summers. The new analytical jurists. *New York University Law Review*, 41:861, 1966.
- [89] M. Sword. To err is both human and non-human. In *University of Missouri-Kansas City Law Review*, 88:211, 2019.
- [90] R. Tricot and B. Sander. Recent developments: The broader consequences of the international court of justice’s advisory opinion on the unilateral declaration of independence in respect of kosovo. *Columbia Journal of Transnational Law*, 49:321-327, 2011.

- [91] Union of Concerned Scientists UCS. UCS Satellite Database. <https://www.ucsusa.org/resources/satellite-database>, 2005.
- [92] The United Nations Office for Outer Space Affairs UNOOSA. Inter-agency meeting on outer space activities: Thirty-eighth session. <https://www.unoosa.org/oosa/en/ourwork/un-space/iam/38th-session.html>, 2018.
- [93] The United Nations Office for Outer Space Affairs UNOOSA. Annual report 2018, 2019.
- [94] United Nations Officer for Outer Space Affairs UNOOSA and Prince Sultan Bin Abdulaziz International Prize for Water PSIPW. Space4water glossary. <https://www.space4water.org/glossary>, n.d.
- [95] F. G. von der Dunk. Sovereignty versus space—public law and private launch in the asian context. *Singapore Journal of International and Comparative Law*, 5:22, 2001.
- [96] F. G. von der Dunk. Outer space law principles and privacy. In D. Leung and R. Purdy, editors, *Evidence from Earth Observation Satellites: Emerging Legal Issues*, page 243–258. Martinus Nijhoff Publishers, 2012.
- [97] F. G. von der Dunk. Legal aspects of navigation. the cases for privacy and liability: An introduction for non-lawyers. <http://mycoordinates.org/legal-aspects-of-navigation/>, 2015. Coodinates magazine, May 2015.
- [98] E. S. Waldrop. Integration of military and civilian space assets: legal and national security implications. *AFL Review*, 55, 157, 2004.
- [99] B. Wang. Us spy satellites at diffraction limit for resolution since 1971, 2019.
- [100] P. Wen and O. Auyezov. Tracking china’s muslim gulag, 2018. 29 November, 2018. Reuters Investigates.
- [101] D. Werner. Self-driving spacecraft? the challenge of verifying ai will work as intended. <https://spacenews.com/self-driving-spacecraft-the-challenge-of-verifying-ai-will-work-as-intended/>. Space News, 2019.
- [102] W. Wiewiórowski. Ai and facial recognition: Challenges and opportunities, 2020. European Data Protection Supervisor, 21 February, 2020.
- [103] S. Yeo. How satellites can help catch disaster insurance fraudsters. <https://psmag.com/environment/new-satellites-can-help-foil-fraud-in-disaster-insurance>, 2018. 3 April.
- [104] Y. Zhao. Revisiting the 1975 registration convention: Time for revision? *Australian Journal of International Law*, 11:106–127, 2004.
- [105] M. Zingler and R. di Marcantonio. Navigating through earth observation knowledge. *The European Space Agency Bulletin*, 96, 1998.